



**You have downloaded a document from  
RE-BUŚ  
repository of the University of Silesia in Katowice**

**Title:** Użycie formalizmu anomalnej dyfuzji oraz funkcji typu Kopuła do badania auto-korelacji i korelacji występujących w słabo przewidywalnych danych fizycznych i finansowych

**Author:** Krzysztof Domino

**Citation style:** Domino Krzysztof. (2014). Użycie formalizmu anomalnej dyfuzji oraz funkcji typu Kopuła do badania auto-korelacji i korelacji występujących w słabo przewidywalnych danych fizycznych i finansowych. Praca doktorska. Katowice : Uniwersytet Śląski

© Korzystanie z tego materiału jest możliwe zgodnie z właściwymi przepisami o dozwolonym użytku lub o innych wyjątkach przewidzianych w przepisach prawa, a korzystanie w szerszym zakresie wymaga uzyskania zgody uprawnionego.



UNIwersYTET ŚLĄSKI  
W KATOWICACH



Biblioteka  
Uniwersytetu Śląskiego



Ministerstwo Nauki  
i Szkolnictwa Wyższego

INSTYTUT FIZYKI im. Augusta Chełkowskiego  
Wydział Matematyki Fizyki i Chemii  
Uniwersytet Śląski w Katowicach

**mgr Krzysztof Domino**

Użycie formalizmu anomalnej dyfuzji oraz funkcji typu  
Kopuła do badania auto-korelacji i korelacji  
występujących w słabo przewidywalnych  
danych fizycznych i finansowych

**Rozprawa Doktorska**

promotor:  
dr hab. Tomasz Błachowicz, prof. Pol. Śl.  
Instytut Fizyki – Centrum Naukowo-Dydaktyczne  
Politechnika Śląska

promotor pomocniczy:  
dr inż. Grzegorz Adamiec  
Instytut Fizyki – Centrum Naukowo-Dydaktyczne  
Politechnika Śląska

Gliwice 2014

Chciałbym gorąco podziękować mojemu promotorowi,

Panu dr hab. prof. Pol. Śl. Tomaszowi Błachowiczowi

Za wszelką pomoc oraz poświęcony czas. Chciałbym też podziękować Panu prof. za wspólne badania naukowe, wspólne publikacje oraz duży wkład w mój rozwój naukowy.

Chciałbym również podziękować mojemu promotorowi pomocniczemu Panu dr inż. Grzegorzowi Adamcowi za poświęcony czas oraz pomoc w przygotowaniu niniejszej dysertacji.

Dziękuję także pracownikom oraz kolegom doktorantom z Instytutu Fizyki Politechniki Śląskiej za otrzymaną pomoc oraz miłą atmosferę. W szczególności chciałbym podziękować Panu dr inż. Grzegorzowi Adamcowi za pomoc w badaniach.

Dziękuję również pracownikom Instytutu Fizyki Uniwersytetu Śląskiego za wszelką pomoc oraz życzliwą atmosferę. W szczególności Panu prof. dr hab. Janowi Składkowskiemu za poświęcony czas oraz pomoc w pierwszych badaniach oraz Panu dr hab. prof. UŚ. Jerzemu Dajce ze współpracę.

Dziękuję także mojej żonie oraz wszystkim, którzy wspierali mnie podczas przygotowania niniejszej dysertacji.

Dedykacja

**Pracę dedykuje mojej żonie oraz córeczce Ali**

## Spis treści

<b>1. Wstęp .....</b>	<b>8</b>
<b>2. Sygnały stochastyczne i deterministyczne w geofizyce i ekonofizyce .....</b>	<b>11</b>
<b>3. Narzędzia i metody badawcze .....</b>	<b>14</b>
<b>3.1. Opis formalizmu dyfuzji anomalnej .....</b>	<b>14</b>
3.1.1. Normalna dyfuzja - podejście klasyczne .....	14
3.1.2. Pojęcie fraktali oraz fraktali statystycznych .....	16
3.1.3. Błądzenie przypadkowe w układach samopodobnych .....	18
<b>3.2. Formalizm anomalnej dyfuzji do badania szeregów czasowych .....</b>	<b>19</b>
<b>3.3. Wykorzystanie formalizmu anomalnej dyfuzji i wykładnika Hursta do modelowania zjawisk słabo przewidywalnych .....</b>	<b>23</b>
<b>3.4. Funkcje Kopuły .....</b>	<b>34</b>
3.4.1. Definicje, zastosowanie i historia .....	34
3.4.2. Podział i rodzaje funkcji Kopuły .....	35
3.4.3. Właściwości Kopuły dwuwymiarowych .....	44
3.4.4. Sposoby estymacji parametrów kształtu .....	51
3.4.5. Sposoby generowania danych .....	52
<b>4. Modelowanie sygnałów geofizycznych i ekonofizycznych .....</b>	<b>55</b>
4.1. Rozkłady prawdopodobieństwa i teoria wartości ekstremalnych .....	55
4.2. Inne rozkłady, dane ekstremalne a auto-korelacje .....	57
4.3. Auto-korelacje a wielowymiarowa statystyka - zastosowanie wykładnika Hursta w procedurze wyboru funkcji Kopuły .....	58
<b>5. Analiza danych empirycznych .....</b>	<b>59</b>
<b>5.1. Badanie izotopów w rdzeniach arktycznych .....</b>	<b>59</b>
5.1.1. Obieg pierwiastków w atmosferze .....	61
5.1.2. Badania auto-korelacji .....	70
5.1.3. Analiza danych z wykorzystaniem funkcji Kopuły .....	75
5.1.4. Analiza danych ekstremalnych .....	89
<b>5.2. Badanie maksymalnych wezbrań sztormowych w Bałtyku.....</b>	<b>92</b>

5.2.1.	Czynniki naturalne wpływające na wahania poziomu morza .....	92
5.2.2.	Badania auto-korelacji .....	94
5.2.3.	Analiza danych z wykorzystaniem funkcji Kopuły .....	102
5.3.	<b>Badanie danych finansowych</b> .....	110
5.3.1.	Badanie danych ekstremalnych .....	110
5.3.2.	Modele typu VaR .....	114
5.3.3.	Modelowanie kursów akcji spółek giełdowych notowanych na GPW w Warszawie .....	115
6.	<b>Podsumowanie pracy</b> .....	123
7.	<b>Literatura</b> .....	128
Dodatki	.....	135
Dodatek A	- transformacja danych, utworzenie szeregów skumulowanych .....	135
Dodatek B	- wyliczenie wykładnika Hursta z użyciem DFA .....	137
Dodatek C	- transformacja danych do przestrzeni $[0, 1]$ oraz utworzenie zmiennej dwuwymiarowej na przestrzeni $[0, 1]^2$ .....	139
Dodatek D	- dopasowanie do danych Kopuł Archimedesowskich .....	142
Dodatek E	- dopasowanie do danych Kopuły Studenta .....	148
Dodatek F	- empiryczne wyznaczenie indeksów górnej oraz dolnej zależności ogonowej dla danych w przestrzeni $[0, 1]^2$ .....	152

## Spis symboli

$x(t)$  – wartość szeregu czasowego w chwili  $t$ ,

GPW – Giełda Papierów Wartościowych,

$$\vec{\nabla} = \vec{i} \frac{\partial}{\partial x} + \vec{j} \frac{\partial}{\partial y} + \vec{k} \frac{\partial}{\partial z},$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \text{operator Laplace'a},$$

$\langle \dots \rangle$  – wartość średnia,

$d_f$  – wymiar fraktalny,

$d_w$  – wymiar fraktalny błędzenia przypadkowego,

$d_s$  – wymiar spektralny,

$N(\varepsilon)$  – ilość jednostkowych klocków o wymiarze  $\varepsilon$ , wypełniających badany obiekt fraktalny,

$\alpha$  - symbol proporcjonalności,

$H$  – wykładnik Hursta,

$H_{threshold}$  – wartość progowa wykładnika Hursta,

DFA – analiza beztrendowa (Detrended Fluctuation Analysis),

$P(t)$  – cena akcji w czasie  $t$ ,

$R(t)$  – wielkość stopy zwrotu z inwestycji w akcje w czasie  $t$ ,

$[ \dots ]$  – funkcja sufitu, najmniejsza liczba całkowita nie mniejsza od argumentu,

$\langle F(\tau)^2 \rangle$  – średnia wariancja beztrendowa dla okna czasowego o długości  $\tau$ ,

WIG – Warszawski Indeks Giełdowy,

## Spis symboli

WIG 20 - Warszawski Indeks Giełdowy dla 20 największych spółek,

$WIG20$  – wartość indeksu WIG 20,

$C$  – funkcja Kopuły,

$f_i(x_i)$  – jednowymiarowa gęstość prawdopodobieństwa zmiennej losowej  $x_i$ ,

$F_i(x_i)$  - jednowymiarowa dystrybuanta zmiennej losowej  $x_i$ ,

$v_i$  – zmienna losowa o rozkładzie jednorodnym na odcinku  $[0,1]$ ,

$[0,1]^n$  – symbol obrazuje  $n$ -wymiarową zmienną losową o rozkładach jednorodnych na odcinku  $[0,1]$ ,

$\Phi_R(\dots)$  – wielowymiarowa dystrybuanta Gaussa o macierzy korelacji  $R$ ,

$C^{Ga}$  – Kopuła Gaussa,

$t_{Rv}$  – wielowymiarowa dystrybuanta Studenta,

$C^T$  – Kopuła Studenta,

$C^G$  – Kopuła Gumbela,

$C^c$  – Kopuła Claytona,

$C^F$  – Kopuła Franka,

$\lambda_u$  – indeks górnej zależności ogonowej funkcji Kopuły,

$\lambda_d$  – indeks dolnej zależności ogonowej funkcji Kopuły,

$\Pr(\dots)$  – prawdopodobieństwo,

$\Pr(\dots | \dots)$  - prawdopodobieństwo warunkowe,

$A_{cm}(t)$  – skumulowany szereg czasowy powstały po transformacji szeregu czasowego  $A(t)$ ,

$K^A$  – koncentracja jonu lub izotopu A zarejestrowana w rdzeniach arktycznych,

$L_{MAX}$  – wartość maksymalnych wezbrań sztormowych zarejestrowanych na Morzu Bałtyckim.



## 1. Wstęp

Analiza szeregów czasowych opisujących zjawiska fizyczne oraz finansowe jest dużym wyzwaniem. W pracy zawarto wyniki badań różnego rodzaju szeregów czasowych opisujących zjawiska geofizyczne oraz finansowe, a także zawarto analizę analogii pomiędzy tymi dwoma rodzajami danych. Celem pracy była ewaluacja ogólnego formalizmu, który może być wykorzystany do analizy korelacji oraz auto-korelacji wyżej wymienionych danych. W ten sposób wyznaczono parametry korelacji oraz auto-korelacji dla następujących danych.

1. Finansowych - kursów akcji notowanych na GPW w Warszawie.
2. Hydrologicznych - maksymalnych wzebrań sztormowych zarejestrowanych w pięciu portach na Morzu Bałtyckim.
3. Klimatycznych - koncentracji izotopów ( $^{18}\text{O}$ ,  $^2\text{D}$  i  $^{10}\text{Be}$ ) i jonów ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ) w rdzeniach lodowych pozyskanych ze stacji badawczej „Taylor Dome” na Antarktydzie.

Do analizy danych użyto następujących narzędzi.

1. Wykładnika Hursta i analizy beztrendowej (Detrended Fluctuation Analysis, DFA) użytych do badania wymiaru fraktalnego oraz anomalnej dyfuzji w przestrzeni badanych szeregów czasowych.
2. Dwuwymiarowych funkcji Kopuły użytych do badania korelacji pomiędzy dwoma szeregami czasowymi. Dodatkowo wykorzystano zależności ogonowe funkcji Kopuły do analizy jednoczesnego występowania zdarzeń ekstremalnych w analizowanych szeregach czasowych.
3. Jednowymiarowych rozkładów prawdopodobieństwa użytych w procedurze dopasowania funkcji Kopuły, w tym rozkładów wartości ekstremalnych. Przeprowadzono ponadto dyskusję stosowania rozkładów wartości ekstremalnych w przypadku danych posiadających auto-korelacje.

Funkcje Kopuły znajdują zastosowanie przy konstrukcji wielowymiarowych rozkładów prawdopodobieństwa szeregów czasowych, po dopasowaniu jednowymiarowych rozkładów brzegowych. Służą one również do analizy korelacji pomiędzy szeregami czasowymi w sposób

bardziej precyzyjny niż przy pomocy jednego współczynnika korelacji, oraz mogą być użyte do oceny ryzyka wystąpienia wartości ekstremalnych. Procedura wyboru właściwej funkcji Kopuły jest skomplikowana. W zaprezentowanej pracy procedura ta została rozszerzona poprzez badanie właściwości fraktalnych oraz auto-korelacji szeregów czasowych. Konceptji tej użyto w modelu oceny ryzyka podczas analizy danych fizycznych - maksymalnych wezbrań sztormowych, oraz danych finansowych - maksymalnych spadków kursów akcji.

Główne tezy pracy można przedstawić następująco:

- istnieje analogiczna dynamika wartości wykładnika Hursta wyliczonych dla niektórych danych geofizycznych i finansowych - historycznej koncentracji jonów pochodzenia morskiego i kursów akcji,
- istnieje możliwość stosowania częściowo stochastycznego i częściowo deterministycznego modelu, uwzględniającego czynniki globalne i lokalne, do analizy szeregów czasowych pochodzących z danych fizycznych oraz finansowych,
- istnieje możliwość zastosowania analogicznego modelu oceny ryzyka wystąpienia maksymalnych wezbrań sztormowych oraz maksymalnych spadków kursów akcji, opracowanego z użyciem funkcji Kopuły i wspieranego dodatkowo przez wykładnik Hursta.

Praca składa się z 5 rozdziałów, podsumowania, spisu literatury oraz dodatków numerycznych. W pierwszym rozdziale pokazano założenia dla modelu analizy danych, wprowadzającego tzw. czynnik globalny, odpowiadający za korelacje i czynniki lokalne o różnej intensywności odpowiadające za właściwości stochastyczne układu fizycznego. Wpływ czynnika globalnego oraz czynników lokalnych odpowiada za mniejsze lub większe skorelowanie danych. Model ten w dalszych rozdziałach zostanie użyty do badania danych złożonych, które posiadają niezerowe i zmienne auto-korelacje.

W drugim rozdziale, w ramach wstępu teoretycznego, omówiono pojęcie normalnej dyfuzji oraz anomalnej dyfuzji, jako błędzenia przypadkowego na fraktalach - układach samopodobnych. Pokazano procedurę wyznaczania wymiarów fraktalnych obiektów oraz wymiaru fraktalnego błędzenia przypadkowego. Wskazano zastosowanie formalizmu anomalnej dyfuzji do badania szeregów czasowych. Omówiono znaczenie wykładnika Hursta, sposoby jego wyznaczania oraz powiązanie wartości wykładnika Hursta z występowaniem auto-korelacji w szeregach czasowych. Następnie zaprezentowano badania autora polegające na zastosowaniu wykładnika Hursta do analizy szeregów czasowych składających się z danych finansowych – kursów akcji notowanych na Giełdzie Papierów Wartościowych (GPW) w Warszawie. W dalszej części rozdziału pokazano formalizm wykorzystany do badania korelacji pomiędzy różnymi szeregami czasowymi, oparty na funkcjach

Kopuły używanych do konstrukcji wielowymiarowych rozkładów prawdopodobieństwa. Przedstawiono własności funkcji Kopuły, które zostały wykorzystane do dalszej analizy.

W trzecim rozdziale opisano narzędzia stosowane do modelowania sygnałów finansowych i geofizycznych. W tym celu omówiono jednowymiarowe, brzegowe rozkłady prawdopodobieństwa użyte wraz z funkcjami Kopuły do analizy danych, które mogą posiadać auto-korelacje. Przeanalizowano również wpływ auto-korelacji danych na procedurę wyboru odpowiedniej funkcji Kopuły.

Czwarty rozdział pracy przedstawia wyniki badań różnego rodzaju, konkretnych szeregów czasowych, to znaczy danych geofizycznych i finansowych. Danymi geofizycznymi była zmienna w czasie koncentracja izotopów ( $^{18}\text{O}$ ,  $^2\text{D}$  i  $^{10}\text{Be}$ ) oraz jonów ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ) w rdzeniach lodowych pozyskanych ze stacji badawczej „Taylor Dome” na Antarktydzie oraz maksymalne wezbrania sztormowe zarejestrowane w różnych portach na Morzu Bałtyckim. Danymi finansowymi były kursy akcji notowane na GPW w Warszawie. W rozdziale stwierdzono podobne wartości wykładnika Hursta dla koncentracji jonów pochodzenia morskiego -  $\text{Na}^+$ ,  $\text{Mg}^{2+}$  oraz danych finansowych. Opracowano tutaj również model oceny ryzyka wystąpienia wartości ekstremalnych stosowany zarówno dla danych geofizycznych - maksymalnych wezbrań sztormowych, jak i finansowych - silnych spadków kursów akcji.

We podsumowaniu pracy przeprowadzono dyskusję analogii pomiędzy dynamiką układów geofizycznych oraz dynamiką danych finansowych. Omówiono praktyczne zastosowanie badanego modelu oceny ryzyka oraz opisano propozycję dalszych badań z tego obszaru. Wskazano również słabe strony takiej analogii, szczególnie założenie, że układ finansowy jest układem otwartym oraz że trudno znaleźć globalne prawa zachowania dla układu finansowego odpowiadające np. prawu zachowania energii lub pędu, które jest spełnione w układach fizycznych.

## 2. Sygnały stochastyczne i deterministyczne w geofizyce i ekonofizyce

W obecnym rozdziale zostanie zaprezentowana analiza sygnałów - szeregów czasowych  $x(t)$  pochodzących z układów geofizycznych i ekonofizycznych. Podczas realizacji pracy badano szeregi czasowe, których dynamika jest częściowo stochastyczna oraz częściowo deterministyczna. W przypadku jednowymiarowym istotnym staje się określenie, w jaki sposób wartość szeregu czasowego  $x(t)$  zależy od wartości przeszłych  $x(t-1)$ ,  $x(t-2)$  ...  $x(t-n)$ . Dzięki temu można odpowiedzieć na pytanie, czy kolejne wartości  $x(t)$  są losowe, a ich wartości są opisane pewnym rozkładem prawdopodobieństwa oraz czy występuje auto-korelacja, która kształtuje wartość  $x(t)$  w zależności od wartości  $x(t-1)$ ,  $x(t-2)$  ...  $x(t-n)$ . Do analizy auto-korelacji można wykorzystać koncepcję anomalnej dyfuzji oraz wykładnika Hursta, co zostanie omówione w kolejnych rozdziałach.

Analogicznie można badać dane dwuwymiarowe złożone z dwóch szeregów czasowych  $(x_i(t), x_j(t))$ . Częściowy determinizm można tutaj powiązać z korelacją pomiędzy danymi. Przy takim podejściu dane w pełni stochastyczne byłyby całkowicie niezależne. Należy tutaj zwrócić uwagę, że jeden współczynnik korelacji nie może opisywać w pełni korelacji pomiędzy dwoma szeregami czasowymi, dlatego do bardziej precyzyjnego opisu szeregów można wykorzystać funkcje Kopuły [1]. Już w tym miejscu można przedstawić postulat, w którym wartości szeregów czasowych są opisane przez czynniki globalne  $G_1, G_2, \dots, G_n$  oraz czynniki lokalne, różne dla różnych sygnałów  $x_i(t)$  oraz  $x_j(t)$ ,  $L_{i1}, L_{i2}, \dots, L_{in}$ ,  $L_{j1}, L_{j2}, \dots, L_{jn}$ . Ponadto, zarówno czynniki lokalne jak również czynniki globalne, mogą być w pełni stochastyczne lub zawierać człon powodujący auto-korelację. Jeżeli założymy, że szeregi czasowe są osobnymi funkcjami pewnych czynników globalnych oraz czynników lokalnych, to możemy zapisać je jako:

$$\begin{aligned} x_i(t) &= f_1(G_1, G_2, \dots, G_n) + f_2(L_{i1}, L_{i2}, \dots, L_{in}), \\ x_j(t) &= f_1(G_1, G_2, \dots, G_n) + f_2(L_{j1}, L_{j2}, \dots, L_{jn}). \end{aligned} \quad (1)$$

Zależność od czynnika globalnego wprowadza korelację pomiędzy szeregami czasowymi, natomiast zależność od czynników lokalnych może, przeciwnie, wprowadzać losowość. Równanie (1) jest

poglądowe, a jego celem jest pokazanie w jaki sposób badając korelacje można odnieść się do czynników lokalnych oraz globalnych kształtujących dane.

Warto jeszcze zwrócić uwagę na różne pary szeregów czasowych  $(x_i(t), x_j(t))$  o różnym stopniu podobieństwa indeksowanego w taki sposób, aby wartość indeksu  $i$  oraz  $j$  reprezentowały podobieństwo danych zdefiniowane jako ogólnie pojętą korelację. Pary szeregów czasowych możemy uszeregować według malejącego podobieństwa od  $(x_1(t), x_2(t))$  poprzez  $(x_1(t), x_i(t))$  aż do  $(x_1(t), x_n(t))$ . Silniejsza korelacja w parze  $(x_1(t), x_2(t))$  świadczy zarówno o wpływie podobnych czynników lokalnych, jak i również o wpływie czynnika globalnego. Z drugiej strony, jeżeli zostanie stwierdzona korelacja dla danych odległych  $(x_1(t), x_n(t))$ , świadczy ona przede wszystkim o wpływie czynnika globalnego. Zauważmy, że badanie auto-korelacji szeregów czasowych oraz porównywanie korelacji różnych szeregów czasowych o różnym stopniu podobieństwa, może być użyte do określenia stopnia stochastyczności czynników globalnych oraz lokalnych, korzystając z następujących założeń:

- podobne wartości wykładnika Hursta, odzwierciedlające auto-korelacje dla różnych szeregów czasowych o różnych korelacjach, świadczą o auto-korelacjach dla czynników globalnych,
- różne wartości wykładnika Hursta dla różnych szeregów czasowych świadczą o różnych auto-korelacjach dla czynników lokalnych.

Przyjrzyjmy się przykładowo danym ekonofizycznym - kursom akcji notowanym na Giełdzie Papierów Wartościowych. Kursy poszczególnych papierów wartościowych  $P_i(t)$  zależą zarówno od czynników globalnych jak również czynników lokalnych [2]. Podstawowym czynnikiem globalnym jest koniunktura makroekonomiczna odzwierciedlona we wzroście gospodarczym oraz zyskowności firm. Duża zyskowność firm zachęca inwestorów do inwestowania w akcje, a co za tym idzie wpływa na wzrost wartości papierów wartościowych. Do grona czynników globalnych należy też zaliczyć np. sytuację polityczną lub relacje międzypaństwowe. Warto w tym miejscu wspomnieć o tzw. ogólnym stanie rynku, czyli np. występowaniu hossy lub bessy. Czynnikiem ten jest związany z kolektywną strategią inwestorów i może być źródłem auto-korelacji, ponieważ np. silne wzrosty na rynku przyciągają kolejnych inwestorów powodując kolejne wzrosty [2] i vice versa.

Czynniki lokalne są związane z poszczególnymi spółkami, jak również sektorami, do których te spółki należą. Czynniki te są np. wyniki finansowe oraz perspektywiczne wyniki finansowe poszczególnych spółek. Dlatego korelacje wewnątrz sektora są ogólnie większe niż pomiędzy sektorami. Jednak do istotnych czynników lokalnych należy też zaliczyć indywidualne strategie

inwestorów, które w pewnych warunkach mogą wpływać w większym stopniu na kursy akcji niż wymienione wyżej czynniki o charakterze racjonalnym [2].

Analizując dane geofizyczne można również wyodrębnić czynniki globalne i lokalne. Czynniki globalne są związane ze wspólnym mechanizmem kształtującym dane, np. są nimi fronty atmosferyczne powodujące wezbrania sztormowe w wielu lokalizacjach, lub wspólny mechanizm obiegu pierwiastków w przyrodzie. Czynniki lokalne wpływają na pojedyncze zmienne np. poprzez lokalny wpływ na wezbrania sztormowe w danej lokalizacji lub lokalny czynnik kształtujący obieg danego izotopu w przyrodzie. W tym miejscu założono, że analogicznie jak dla danych finansowych, również dla danych geofizycznych zarówno czynniki globalne, jak też lokalne, mogą mieć „wewnętrzną strukturę” odzwierciedloną w auto-korelacjach danych geofizycznych.

### 3. Narzędzia i metody badawcze

#### 3.1. Opis formalizmu anomalnej dyfuzji

##### 3.1.1 Normalna dyfuzja – podejście klasyczne

W podrozdziale tym zostanie omówiona normalna dyfuzja, będąca wstępem do uogólnienia - anomalnej dyfuzji, której formalizm zostanie wykorzystany do wyznaczenia wykładnika Hursta oraz badania auto-korelacji przedstawionych szeregów czasowych.

Aby wyprowadzić równanie dyfuzji zapiszmy gęstość materii w punkcie  $r$  oraz w czasie  $t$  jako  $\phi(r, t)$ . Rozważmy następnie równanie ciągłości:

$$\frac{\partial \phi}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0, \quad (2)$$

gdzie  $\vec{J}$  to strumień materii. Zastosujmy pierwsze prawo Ficka, które mówi, że strumień materii jest proporcjonalny do lokalnego gradientu gęstości materii:

$$\vec{J} = -D(\phi) \vec{\nabla} \cdot \phi, \quad (3)$$

Otrzymujemy zatem

$$\frac{\partial \phi}{\partial t} = \vec{\nabla} \cdot (D(\phi) \vec{\nabla} \cdot \phi). \quad (4)$$

Jeśli wielkość  $D$  jest stała, wtedy otrzymujemy klasyczne równanie dyfuzji,

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi. \quad (5)$$

Rozważmy równanie dyfuzji w jednym wymiarze:

$$\frac{\partial \phi}{\partial t} - D \frac{\partial^2 \phi}{\partial x^2} = 0 \quad (6)$$

z warunkiem początkowym  $\phi(x, 0) = \phi_0(x)$ . Wyliczamy funkcję Greena:

$$G(x, \xi, t) = \frac{1}{2\sqrt{\pi Dt}} e^{-\frac{(x-\xi)^2}{4Dt}}. \quad (7)$$

Równanie genetyczne jest opisane poprzez:

$$\phi(x, t) = \int_{-\infty}^{\infty} G(x, \xi, t) \phi_0(\xi) d\xi. \quad (8)$$

Jeżeli założymy, że wszystkie cząstki znajdowały się w czasie  $t = 0$  w punkcie  $x_0$  możemy przedstawić warunki początkowe jako  $\phi_0(x) = \delta(x - x_0)$ . Wtedy otrzymujemy następujące rozwiązanie równania dyfuzji

$$\phi(x, t) = \frac{1}{2\sqrt{\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}}. \quad (9)$$

Jest to równanie opisujące krzywą Gaussa z wariancją równą  $2Dt$  oraz odchyleniem standardowym  $\sigma = \sqrt{2Dt}$ .

Ogólnie średnie kwadratowe przemieszczenie cząstek podlegających normalnej dyfuzji, zapisane tutaj jako  $\langle r^2(t) \rangle$ , jest proporcjonalne do czasu a jego pierwiastek jest proporcjonalny do pierwiastka z czasu, zatem:

$$\langle r^2(t) \rangle \propto t, \quad (10)$$

$$\sqrt{\langle r^2(t) \rangle} \propto \sqrt{t}. \quad (11)$$

W celu zobrazowania dyfuzji rozpatrzmy jednowymiarowe błądzenie przypadkowe. Ciało rozpoczyna ruch w punkcie  $x_0 = 0$ , w jednostkowym czasie wykonuje ruch w którąś ze stron z prawdopodobieństwem  $\frac{1}{2}$ . Zmianę pozycji ciała opisują zmienne losowe  $Z_1, Z_2, \dots, Z_n$  równe  $\pm 1$



z prawdopodobieństwem  $\frac{1}{2}$ . Przebyta droga po  $n$  krokach jest sumą zmiennych losowych  $Z_i$ . Wtedy wartość średnia z przebytej drogi -  $\langle x(n) \rangle$  jest równa:

$$\langle x(n) \rangle = \left\langle \sum_{i=1}^n Z_i \right\rangle = 0. \quad (12)$$

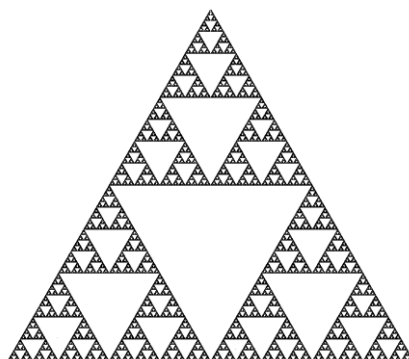
Rozważmy teraz średnią kwadratową z przebytej drogi -  $\langle x^2(n) \rangle$ :

$$\langle x^2(n) \rangle = \left\langle \left( \sum_{i=1}^n Z_i \right)^2 \right\rangle = \sum_{i=1}^n \langle Z_i^2 \rangle + 2 \sum_{i>j}^n \langle Z_i \cdot Z_j \rangle = n + 2 \sum_{i>j}^n \langle Z_i \cdot Z_j \rangle. \quad (13)$$

Widać więc, że w przypadku, gdy kolejne kroki są niezależne do siebie, a dyfuzja odbywa się w przestrzeni euklidesowej, człon  $\sum_{i>j}^n \langle Z_i \cdot Z_j \rangle$  jest równy zero. Otrzymujemy zatem  $\langle x^2(n) \rangle = n$ , co oznacza, jak pokazano wcześniej, że średnia droga kwadratowa jest proporcjonalna do ilości kroków  $n$  lub czasu  $t$  (równość wynika z faktu, że założono wykonanie jednostkowego kroku w jednostkowym czasie). Jeśli jednak kolejne kroki zależałyby od siebie w/w warunek nie byłby spełniony [3] i  $\sum_{i>j}^n \langle Z_i \cdot Z_j \rangle \neq 0$ . Wtedy mówimy o anomalnej dyfuzji, ponieważ średnia droga kwadratowa nie jest już proporcjonalna do czasu. W przypadku anomalnej dyfuzji klasyczne równanie dyfuzji nie jest spełnione, ponieważ dyfuzja nie odbywa się w przestrzeni euklidesowej - niecała przestrzeń znajdująca się w najbliższym otoczeniu ciała należy do przestrzeni dyfuzji [3]. Przestrzeń tą opisuje się za pomocą pojęcia fraktali.

### 3.1.2 Pojęcie fraktali oraz fraktali statystycznych

Fraktale są matematycznymi obiektami, dla których wymiar Hausdorffa nie jest liczbą całkowitą. Podstawowymi cechami fraktali jest samopodobieństwo w każdej skali oraz fakt, że nie są one gładkie w żadnym punkcie. Pojęcie fraktala zostało po raz pierwszy użyte przez Mandelbrota [4] w 1977 roku, jednak badania obiektów typu fraktalnego prowadzone były przez matematyków już dużo wcześniej. Przykładem obiektu fraktalnego jest Trójkąt Sierpińskiego - rysunek 1.



Rysunek 1. Trójkąt Sierpińskiego.

Na rysunku nr 1 można zauważyć, że biała przestrzeń nie należy do Trójkąta Sierpińskiego, jest to istotne, ponieważ przestrzeń ta jest niedostępna dla dyfuzji. Jeżeli powiększymy każdy z wymiarów liniowych 2 razy uogólniona powierzchnia trójkąta wzrośnie 3 razy. W celu wyliczenia wymiaru fraktalnego  $d_f$  figury korzystamy z zależności:

$$3 = 2^{d_f}, \quad (14)$$

$$d_f = \frac{\ln 3}{\ln 2} = 1.585. \quad (15)$$

Wymiar fraktalny jest mniejszy niż 2, mimo że badamy obiekt w euklidesowej przestrzeni dwuwymiarowej.

Trójkąt Sierpińskiego jest przykładem matematycznego fraktala, obiektu samopodobnego w każdej skali długości. W przyrodzie występują statystyczne fraktale, obiekty, które mają cechy samopodobieństwa w sensie bardziej statystycznym niż deterministycznym. Jako przykład można podać przestrzenne ułożenie wielu polimerów, powierzchnię chmur, powierzchnię księżyca czy płatki śniegu. Aby wyznaczyć wymiar fraktalny statystycznego fraktala należy zbadać jak jego ogólnie zrozumiana „masa”  $M$  skaluje się z wymiarem liniowym  $l$  [3]:

$$M \propto l^{d_f}. \quad (16)$$

Niektóre obiekty fizyczne można opisać przy pomocy narzędzi matematycznych używanych do opisu układów fraktalnych. Badania takie przeprowadzono np. dla jaskiń gipsowych znajdujących się w Zachodniej Ukrainie [5, 6]. W celu wyznaczenia, w jaki sposób „masa” obiektu skaluje się z wymiarem liniowym można wyliczyć (pojemnościowy) wymiar fraktalny  $d_f$  poprzez zliczanie jednostkowych klocków o wymiarze równym  $\varepsilon$  wypełniających obiekt. W przypadku

jednowymiarowego obiektu liczba klocków  $N(\varepsilon)$  jest powiązana z wymiarem liniowym obiektu  $L$  poprzez następujące równanie:

$$L = N(\varepsilon) \cdot \varepsilon, \quad (17)$$

a w przypadku obiektu samopodobnego mamy [5, 6]:

$$L^{d_f} = N(\varepsilon) \cdot \varepsilon^{d_f}, \quad (18)$$

dzięki czemu możemy wyliczyć wymiar fraktalny  $d_f$ . Logarytmując obustronnie (18) oraz zakładając  $d_f > 1$  otrzymujemy [5, 6]:

$$d_f = \frac{\ln N(\varepsilon)}{\ln L + \ln\left(\frac{1}{\varepsilon}\right)} \approx \frac{\ln N(\varepsilon)}{\ln\left(\frac{1}{\varepsilon}\right)}. \quad (19)$$

Wykonanie regresji liniowej  $\ln N(\varepsilon)$  względem  $\ln\left(\frac{1}{\varepsilon}\right)$  pozwoliło wyliczyć wartość  $d_f$ , która dla jaskiń znajdujących się na Zachodniej Ukrainie znajdowała się w przedziale  $1.7 < d_f < 1.8$  [5, 6]. Badane obiekty traktowano z dużym przybliżeniem jako obiekty dwuwymiarowe, dlatego wynik  $d_f < 2$  świadczy, że są one fraktalami statystycznymi.

### 3.1.3 Błądzenie przypadkowe w układach samopodobnych

Przeanalizujmy obecnie zjawisko transportu w fizycznych, niekryształicznych układach złożonych - np. transport w popękanych i porowatych skałach, przewodnictwo superjonowych przewodników, czy transport na klasterach perkolacyjnych. Transport nie odbywa się tu w oparciu o formalizm normalnej dyfuzji i błądzenia przypadkowego jak w klasycznych układach krystalicznych [3]. Przypomnijmy, że proces błądzenia przypadkowego można opisać przez następujące po sobie kroki, które z równym prawdopodobieństwem mogą wystąpić w każdym kierunku. Aby opisać błądzenie przypadkowe w układzie złożonym, należy zauważyć, że niecałe najbliższe otoczenie obiektu podlegającego błądzeniu należy do przestrzeni dyfuzji. Z fizycznego punktu widzenia na każdej skali długości pojawiają się pułapki utrudniające i spowalniające dyfuzję.

Jak już powiedziano, matematycznymi figurami, które są samopodobne oraz zajmują tylko część przestrzeni euklidesowej są fraktale. Błądzenie przypadkowe, odbywające się na fraktalu, to dobry przykład do matematycznego opisanie anomalnej dyfuzji. Wyobraźmy sobie zatem

błądzenie przypadkowe na Trójkącie Sierpińskiego, który został już przedstawiony na rysunku 1. Wykazano [3], że średnie kwadratowe przemieszczenie  $\langle r^2(t) \rangle$ , na obiekcie fraktalnym, opisane jest równaniem

$$\langle r^2(t) \rangle \propto t^{\frac{2}{d_w}}, \quad (20)$$

gdzie  $t$  to czas dyfuzji a wartość  $d_w$  to wymiar fraktalny błędzenia przypadkowego, który dla anomalnej dyfuzji jest różny od 2. W naszym przypadku  $d_w = \frac{\ln 5}{\ln 2} = 2.322$  [3]. Przypomnijmy, że wymiar fraktalny Trójkąta Sierpińskiego wynosi  $d_f = \frac{\ln 3}{\ln 2} = 1.585$ . Zależność pomiędzy wymiarem fraktalnym dyfuzji oraz wymiarem fraktalnym obiektu jest złożona, ale można ją opisać badając spektralny wymiar fraktalny oraz tzw. sprzężenie Aleksandra - Orbacha [3]. Wymiar spektralny  $d_s = \frac{2d_f}{d_w}$  opisuje skalowanie gęstości stanów w układzie złożonym oraz ilości punktów  $S(t)$  napotkanych podczas błędzenia przypadkowego [3]:

$$S(t) \propto t^{\frac{d_s}{2}}. \quad (21)$$

Jednakże dokładna wartość  $d_s \approx \frac{4}{3} = 1.33$  sugerowana przez sprzężenie Aleksandra – Orbacha [7] nie jest uniwersalna [3, 8]. Wyznaczywszy wymiary fraktalne błędzenia przypadkowego w jaskiniach znajdujących się na Zachodniej Ukrainie [9] porównano je z wymiarami fraktalnymi jaskiń [5, 6]. W ten sposób otrzymano  $1.5 \leq d_s \leq 1.64$ , wartości te są trochę większe niż sugerowane  $d_s \approx \frac{4}{3}$ . Należy jednak w tym miejscu zaznaczyć, że badane jaskinie są fraktalami statystycznymi natomiast sprzężenie Aleksandra - Orbacha było badane dla fraktali matematycznych, np. dla błędzenia przypadkowego na Trójkącie Sierpińskiego otrzymuje się  $d_s = 1.365 \approx \frac{4}{3}$ .

### 3.2. Formalizm anomalnej dyfuzji do badania szeregów czasowych

Rozważmy błędzenia przypadkowe podczas anomalnej dyfuzji. Odwrotność wymiaru fraktalnego błędzenia przypadkowego nazywamy wykładnikiem Hursta  $H = \frac{1}{d_w}$  [10, 11]. Dlatego pierwiastek ze średniego kwadratowego przemieszczenia można opisać równaniem:

$$\sqrt{\langle r^2(t) \rangle} \propto t^{\frac{1}{d_w}} = t^H. \quad (22)$$

Wielkość tę można porównać ze średnim kwadratowym przemieszczeniem dla ruchów Browna, czyli  $\sqrt{\langle r^2(t) \rangle} \propto t^{\frac{1}{2}}$ . Zauważmy, że dla błędzenia przypadkowego na fraktalach mamy  $H < \frac{1}{2}$ , a dyfuzja odbywa się wolniej niż w przypadku ruchów Browna. Uogólniając formalizm można otrzymać również przypadek „superdyfuzji”, gdzie  $H > \frac{1}{2}$ . Tutaj dyfuzja odbywa się szybciej niż w przypadku ruchów Browna, z powodu istnienia długozasięgowych dodatnich auto-korelacji pomiędzy kolejnymi krokami. Podsumowując możemy stwierdzić, że wykładnik Hursta jest wygodnym sposobem opisu anomalnej dyfuzji [11, 12], tak więc:

- jeżeli  $H > \frac{1}{2}$  układ posiada długozasięgowe dodatnie auto-korelacje,
- jeżeli  $H = \frac{1}{2}$  kolejne kroki nie zależą od poprzednich,
- jeżeli  $H < \frac{1}{2}$  układ posiada ujemne auto-korelacje.

Błądzenie przypadkowe podczas anomalnej dyfuzji jest ponadto samopodobne w sensie statystycznym [12]:

$$\langle r^2(at) \rangle = a^{2H} \langle r^2(t) \rangle. \quad (23)$$

Równanie (23) spełnione jest zarówno dla  $H = \frac{1}{2}$  jak również dla  $H \neq \frac{1}{2}$ , dlatego znajduje zastosowanie w procedurze wyznaczania wykładnika Hursta oraz ocenie auto-korelacji badanych szeregów czasowych.

Jak pokazano w poprzednim paragrafie, wykładnik Hursta może być użyty do badania auto-korelacji szeregów czasowych. Jest wiele metod wyznaczania wykładnika Hursta [13]. Ogólnie polegają one na ustaleniu okna czasowego o zmiennej długości (oznaczonej przez  $\tau$ ), w którym badamy szereg czasowy. Następnie, należy określić jak „ilość fluktuacji” zarejestrowana w tym oknie czasowym skaluje się z jego długością, aby zastosować wzory (22, 23). Najbardziej znane metody wyliczania wykładnika Hursta to:

- oryginalna metoda  $R/S$  [14],
- metoda analizy beztrendowej (DFA) [15],
- inne metody (np. bayesowska) [16].

Poniżej zostanie omówiona metoda DFA oraz jej zastosowanie do analizy szeregów czasowych.

Analizę DFA rozpoczynamy od określenie okna czasowego o długości  $T$ , w którym będzie badany szereg czasowy  $x(t)$ , przy czym:

- okno może być równe całkowitej długości szeregu czasowego, przeprowadzamy wtedy globalną DFA opisującą globalne własności auto-korelacji,
- może ono również pokrywać tylko część szeregu czasowego i być użyte do analizy lokalnych właściwości tej części szeregu czasowego - wtedy przeprowadzamy lokalną analizę DFA.

Mając wyznaczone okno czasowe o długości  $T$  obliczamy wariację beztrendową  $F(T)^2$  - według wzoru (24). Wyliczenie wariacji beztrendowej polega na obliczeniu średniej kwadratowej różnicy pomiędzy wartością szeregu czasowego  $x(t)$  oraz wartością przewidywaną  $r(t, T)$ . Gdzie  $r(t, T)$  to wartość przewidywana przez trend liniowy dopasowany do szeregu  $x(t)$  w oknie czasowym o długości  $T$ , np. metodą najmniejszych kwadratów:

$$F(T)^2 = \frac{\sum_{t=1}^T (x(t) - r(t, T))^2}{T}. \quad (24)$$

Metoda DFA jest często stosowana do analizy danych finansowych [1, 12, 17 - 25]. Użycie wariacji beztrendowej można uzasadnić faktem, iż badanie trendów liniowych znajduje szerokie zastosowanie w analizie danych finansowych (np. kursów akcji) [26], ponieważ inwestorzy często kierują się obserwacją takich trendów liniowych przy podejmowania decyzji inwestycyjnych, co może wpłynąć na kontynuację trendu liniowego, na zasadzie autosugestii. Trend liniowy staje się wtedy pewnego rodzaju lokalnym stanem równowagi, która może ujawniać się w różnych skalach czasowych, gdyż poszczególni inwestorzy stosują również różne okresy inwestycyjne.

W następnym kroku należy podzielić okno czasowe o długości  $T$  na  $n$  niepokrywających się okien czasowych o długości  $\tau$ , takich, że  $\tau = \frac{T}{n}$ . Dla każdego okna czasowego o długości  $\tau$  wyliczamy wariację beztrendową według wzoru:

$$F(\tau)^2 = \frac{\sum_{t=k}^{\tau+k-1} (x(t) - r(t, \tau))^2}{\tau}, \quad (25)$$

gdzie wielkość  $k$  odpowiada początkowi badanego okna czasowego. Funkcję regresji liniowej  $r(t, \tau)$  dopasowuje się dla każdego z okien czasowych osobno. Następnie należy wyliczyć średnią wariację beztrendową po wszystkich  $n$  oknach czasowych:

$$\langle F(\tau)^2 \rangle = \frac{\sum_{i=1}^n F(\tau)^2}{n}. \quad (26)$$

W tym miejscu warto zwrócić uwagę, że wariancja beztrendowa mierzy „fluktuację” w oknie czasowym o długości  $\tau$ . W następnym kroku stosujemy standardową relację DFA wywodzącą się ze wzoru (22):

$$\langle F(\tau)^2 \rangle \propto \tau^{2H}, \quad (27)$$

lub

$$\sqrt{\langle F(\tau)^2 \rangle} \propto \tau^H. \quad (28)$$

Wykładnik Hursta można ostatecznie wyliczyć jako nachylenie prostej  $\ln \sqrt{\langle F(\tau)^2 \rangle}$  w funkcji  $\ln \tau$  stosując metodę regresji liniowej - np. metodę najmniejszych kwadratów. Należy jednak pamiętać, że metodę DFA stosuje się do analizowania szeregów czasowych typu:

$$x(i) = x(i-1) + \Delta x(i), \quad (29)$$

gdzie  $\Delta x(i)$  odpowiada zmianie wartości szeregu czasowego, w której zawarty jest człon stochastyczny  $W$  oraz człon deterministyczny opisujący auto-korelacje  $k(x(1), x(2), \dots, x(i-1))$ ,

$$\Delta x(i) = W + k(x(1), x(2), \dots, x(i-1)). \quad (30)$$

Tego typu danymi są niektóre dane finansowe, np. kursy akcji  $P(t)$ . Zmiana kursów akcji  $\Delta P(t) = P(t) - P(t-1)$  jest częściowo stochastyczna oraz częściowo deterministyczna [12]:

$$\Delta P(i) = W + k(\Delta P(1), \Delta P(2), \dots, \Delta P(i-1)). \quad (31)$$

Jeżeli badamy bezpośrednio dane typu  $\Delta x(i)$  - np. stopy zwrotu  $\Delta P(i)$  lub niektóre dane geofizyczne, należy je przetransformować do przestrzeni  $x(i)$  [12], stosując następujące równania rekurencyjne:

$$x(1) = x_0 + \Delta x(1), \quad (32)$$

$$x(i) = x(i-1) + \Delta x, \quad (33)$$

a następnie można zastosować standardową procedurę DFA. Omówiona w tym rozdziale procedura wyznaczania wykładnika Hursta zostanie wykorzystana w następnych rozdziałach do analizy autokorelacji.

### **3.3. Wykorzystanie formalizmu anomalnej dyfuzji i wykładnika Hursta do modelowania zjawisk słabo przewidywalnych**

W rozdziale tym zostanie omówione wykorzystanie wykładnika Hursta do analizy słabo przewidywalnych zjawisk na przykładzie analizy danych finansowych - kursów akcji oraz indeksów giełdowych. Jak już wspomniano wcześniej dane finansowe można badać jako szeregi czasowe, np. szeregi czasowe kursów akcji  $P(t)$  wyrażonych w jednostkach monetarnych. Wielkość  $t$  to czas, jakiemu odpowiada dany kurs akcji. Załóżmy, że badamy kursy akcji notowane w sposób ciągły, w ten sposób notowane są na GPW średnie oraz duże spółki [2]. Badania mogą być prowadzone w różnych skalach czasowych [2, 12], można badać: dzienne, minutowe, sekundowe kursy akcji lub kursy z każdej transakcji. Dodatkowo w przypadku dziennych kursów akcji mamy do wyboru kilka możliwości: kursy otwarcia, maksymalne, minimalne lub kursy zamknięcia.

Przeanalizujmy teraz szczegółowo procedurę wyznaczania kursów akcji podczas notowań ciągłych. Kurs otwarcia oraz kurs zamknięcia wyznaczone są w sposób „statyczny” - w fazie przed otwarciem lub przed zamknięciem, gdzie najpierw zbiera się wszystkie zlecenie kupna i sprzedaży, a następnie wybiera się optymalny (równowagowy) kurs. Szczegóły funkcjonowania GPW w Warszawie przedstawiono w [2]. Po wyznaczeniu kursu otwarcia odbywa się faza notowań ciągłych. Proces przyjmuje postać dynamiczną, kursy akcji są wyznaczone na bieżąco podczas realizacji kolejnych zleceń kupna oraz sprzedaży - jest to kurs chwilowy odpowiadający każdej transakcji [2]. Spośród kursów chwilowych w danym dniu wybiera się kurs maksymalny oraz kurs minimalny. Ponadto dzieląc okres notowań ciągłych na przedziały (np. minutowe lub sekundowe) można wyznaczyć kurs akcji w różnych skalach czasowych - np. co minutę lub co sekundę.

Aby pokazać w jaki sposób wykładnik Hursta ( $H$ ) oraz analiza beztrendowa (DFA) mogą być wykorzystane do analizy oraz przewidywania kursów papierów wartościowych należy przyjrzeć się analogii pomiędzy układem finansowym i dynamicznym układem złożonym [12, 17 - 20]. Analogia ta jest związana z poszukiwaniem log-periodycznych oscylacji układu finansowego przed osiągnięciem stanu, analogicznego do przejścia fazowego, odzwierciedlającego załamanie się rynku finansowego [27 - 32]. Wartość wykładnika Hursta równa  $H = 0.5$  odzwierciedla stochastyczne zachowanie się kursów akcji według równania (34):



$$P(t) = P(t - 1) + W(t), \quad (34)$$

gdzie  $W(t)$  to czynnik w pełni stochastyczny. W równaniu (34) można ująć również niewielki dryf deterministyczny [12] związany np. z ogólnym rozwojem gospodarczym i wzrostem Produktu Krajowego Brutto (PKB). Duże zmiany, a w szczególności spadki kursów akcji zaobserwowane na rynkach finansowych [26] nie są dobrze opisane przez model stochastyczny [2, 12] przyjęty w równaniu (34), a co a tym idzie straty wynikające z inwestowanie w akcje, podczas silnego załamania rynku, są znacznie większe niż te przewidywane przez model stochastyczny, zakładający że równość  $H = 0.5$  jest spełniona dla całej historii transakcji. Sytuacja ta może być dość problematyczna dla indywidualnych oraz instytucjonalnych posiadaczy papierów wartościowych, a w skrajnym przypadku może doprowadzić do bankructwa instytucji. Niestety, historia zna wiele przypadków takich bankructw. Z drugiej strony, jeżeli wartość wykładnika Hursta spełnia nierówność  $H < 0.5$  zachowanie kursów akcji nie jest w pełni stochastyczne, ponieważ występują anty-korelacje pomiędzy kolejnymi kursami akcji, co stwierdzono wielokrotnie przed zbliżającymi się załamaniem rynku, czyli dużymi spadkami kursów akcji lub wartości indeksów [12, 19, 20].

W pracy [19] badano indeks WIG odzwierciedlający kursy akcji wszystkich spółek notowanych na GPW w Warszawie. Indeks WIG (Warszawski Indeks Giełdowy) wyraża wartość spółek notowanych na rynku podstawowym na GPW w Warszawie w stosunku do ich wartości z początku notowań indeksu, 16 kwietnia 1991 roku, kiedy jego wartość wynosiła 1000 pkt. Przy wyliczaniu wartości indeksu uwzględnia się głównie ceny akcji -  $P_i(t)$  - notowane na rynku podstawowym GPW w Warszawie, gdzie indeks  $i$  numeruje kolejne spółki, których akcje są notowane. W indeksie WIG uwzględnia się również dochody z dywidendy oraz praw poboru związanych z wyżej wymienionymi papierami wartościowymi. Jednakże wpływ tych ostatnich na wartość indeksu jest znacznie mniejszy niż wpływ kursów akcji [33]. W pracy [19] analizie poddano załamania indeksu WIG ujęte w tabeli 1.

Tabela 1. Załamania indeksu WIG na GPW w Warszawie omówione w pracy [19].

Data	Procentowy spadek wartości	Czas trwania załamania
17.03.1994	65%	41 sesji
22.07.1998	39%	30 sesji
15.05.2006	21%	24 sesje

W [19] przedstawiono również pewne warunki, jakie powinien spełnić lokalny wykładnik Hursta  $H$ , wyliczony dla indeksu WIG, aby można było stwierdzić wystąpienie sygnału anty-korelacyjnego. Ogólnie sprowadzają się one do warunku  $H \ll 0.5$ . Okazało się, że sygnał anty-korelacji został zarejestrowany przed każdym ze spadków indeksu WIG przedstawionym w tabeli 1. Badania zakończono 27 lipca 2007 roku, jednakże w końcowym etapie badań stwierdzono pojawienie się kolejnego sygnału anty-korelacji sugerującego zbliżające się załamanie indeksu WIG. Przewidywania okazały się słuszne, ponieważ najwyższa wartość indeksu WIG (67 568.51 pkt.) została zarejestrowana 6 lipca 2007 roku a następnie nastąpił duży spadek wartości indeksu - powiązany w późniejszej fazie z ostatnim światowym kryzysem gospodarczym. W sumie, w pracy [19] pokazano metodologię, która może być udoskonalana w dalszym badaniu zachowania się rynków papierów wartościowych, w tym GPW w Warszawie.

Należy w tym miejscu zwrócić uwagę, że indeks WIG jest obliczany jako średnia ważona wartości kursów akcji [33]. Dlatego analiza indeksu WIG jest w pewnym sensie uśrednioną analizą kursów akcji - związaną z czynnikami globalnymi kształtującymi obraz całego rynku. W dalszej części dysertacji przedstawię zastosowanie metodologii badania wykładnika Hursta do analizy poszczególnych kursów akcji notowanych na GPW, co przybliży nas do badania czynników lokalnych kształtujących zachowanie się poszczególnych spółek.

W tym miejscu omówione zostaną wyniki pracy autora opublikowane w [24], w której użyto wykładnika Hursta do badania zmian kursów akcji  $P_i(t)$  oraz stóp zwrotu  $R_i(t)$  wyznaczonych dla akcji notowanych na GPW w Warszawie. W pracy stopy zwrotu  $R(t)$  były traktowane jako procentowe zmiany kursów akcji według równania (35):

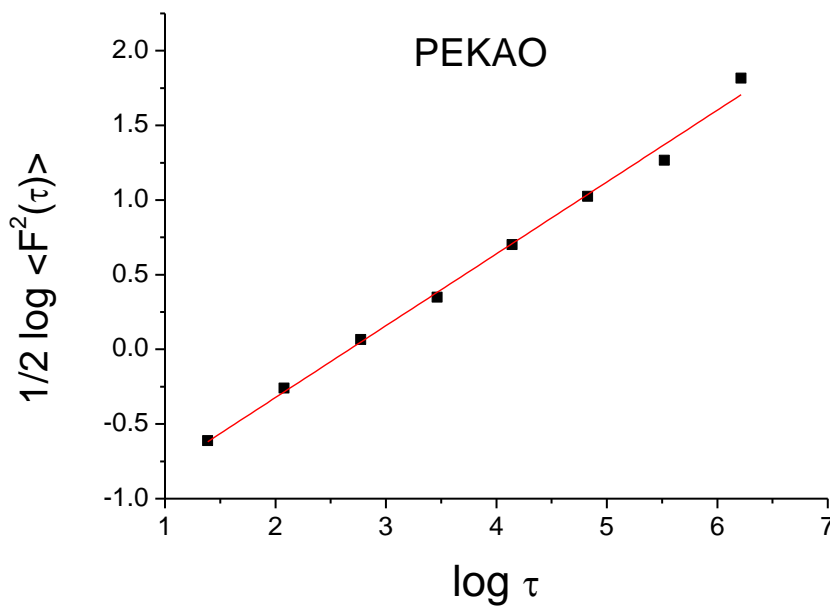
$$R(t) = \frac{\Delta P(t)}{P(t)} = \frac{P(t) - P(t-1)}{P(t)}. \quad (35)$$

Nowatorskie podejście pracy [24] polegało na tym, że wykładnik Hursta był wyznaczany osobno dla każdej z badanych spółek, a badano w sumie 126 spółek notowanych na GPW w Warszawie w okresie 1991 - 2008. Wykładnik Hursta okazał się istotnie statystycznym wyznacznikiem zmian kursów akcji oraz stóp zwrotu. W celu wyznaczenia wykładnika  $H$  użyto lokalnej analizy DFA, w której zastosowano  $N = 500$  punktowe okno obserwacyjne. Podczas badań uwzględniono 2 wartości kursów akcji dziennie - kurs otwarcia oraz kurs zamknięcia. Kursy te były wyznaczane z użyciem podobnej procedury, w sposób statyczny, po przyjęciu wszystkich zleceń w fazie „przed otwarciem” oraz w fazie „przed zamknięciem”. Dlatego kurs otwarcia oraz kurs zamknięcia mogą być ze sobą porównane i dane składające się z szeregu czasowego kolejno następujących po sobie kursów

otwarcia oraz zamknięcia mogą być analizowane w celu wyznaczenia auto-korelacji. Następnie wyznaczono wariancję beztrendową  $F(\tau)^2$  dla różnych okien czasowych o długości  $\tau$  takich, że  $\tau = \left\lfloor \frac{N}{2^k} \right\rfloor$ , gdzie  $k = 0, 1, 2, \dots, 7$ . Zaprezentowany sposób wyboru  $\tau$  powoduje, że punkty są równomiernie rozłożone na osi  $\log \tau$ , co ma wpływ na dokładność dopasowania linii prostej. W celu wyznaczenia wykładnika Hursta wykorzystano relację:

$$\langle F(\tau)^2 \rangle \propto \tau^{2H} \quad (36)$$

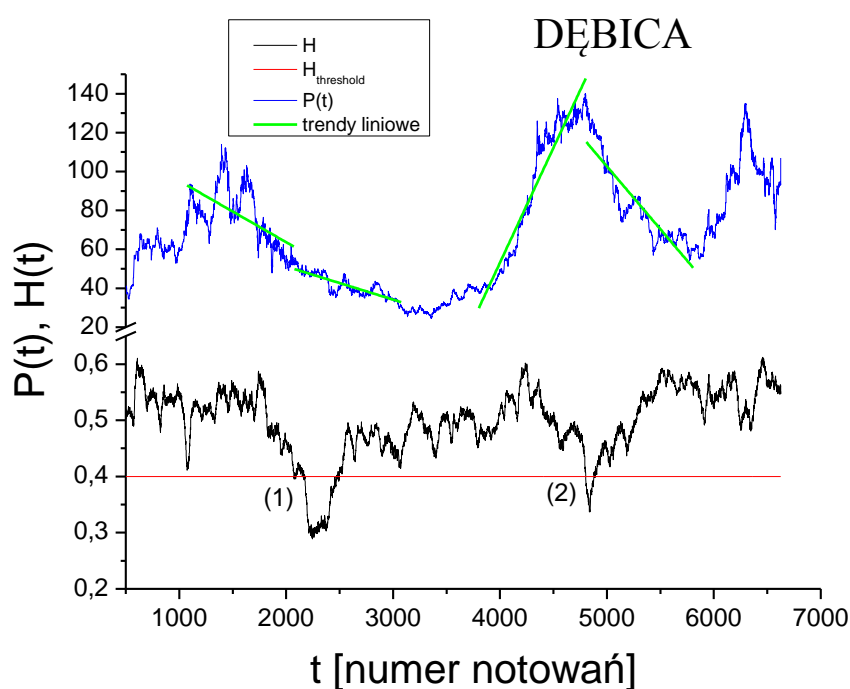
oraz wyznaczono współczynnik nachylenia prostej dopasowany do danych  $\frac{\log \langle F^2(\tau) \rangle}{2}$  w funkcji  $\log \tau$ . Na rysunku 2 przedstawiono przykładowe dopasowanie linii prostej do w/w danych w celu wyliczenia wykładnika Hursta dla spółki PEKAO - szczegóły przedstawiono w pracach [24, 25].



Rysunek 2. Dopasowanie linii prostej metodą regresji liniowej, oraz wyznaczenie jej nachylenia równego wykładnikowi Hursta, analizę przeprowadzono dla danych  $\frac{\log \langle F^2(\tau) \rangle}{2}$  w funkcji  $\log \tau$ .

Wartości wykładnika Hursta  $H$  zostały użyte do poszukiwania anty-korelacji występujących w szeregach czasowych badanych kursów akcji -  $P_i(t)$ . Postawiono hipotezę badawczą, że spadek wykładnika Hursta poniżej wartości progowej  $H_{threshold} < 0.5$  jest wyznacznikiem zbliżającej się zmiany linowego trendu - określonego jako nachylenie prostej dopasowanej do  $P_i(t)$  w zadanym zakresie  $t$ . Było to nowatorskie podejście, ponieważ do tej pory badano wartości indeksów a nie

poszczególnych kursów akcji, a wykładnik Hursta był rozpatrywany jako sygnał zbliżającego się krachu, a nie zmiany trendu. Na rysunku 3 przedstawiono przykład analizy kursów akcji, trendów oraz wykładnika Hursta dla spółki Dębica. Na skali poziomej przedstawiono czas  $t$  związany z kolejnymi transakcjami. Ponieważ badamy dwa kursy akcji dziennie, dwa punkty (kolejne wartości  $t$ ) odpowiadają jednemu dniu transakcji (wartość  $t = 1$  odpowiada pierwszemu dniu transakcji na GPW - w przypadku spółki Dębica to 24 listopada 1994 r. Na wykresie umieszczono dane dla  $t \geq 500$ , ponieważ dopiero dla  $t = 500$  było możliwe wyliczenie pierwszej wartości wykładnika  $H$ . Spadek wykładnika Hursta poniżej wartości progowej  $H_{threshold} = 0.4$  wystąpił w dwóch punktach (1) oraz (2) - dla  $t_1 = 2072$  oraz  $t_2 = 4802$ . Założono, że jest to związane z wystąpieniem w tych punktach sygnału anty-korelacji. Następnie zbadano trendy liniowe o długości  $|t| = 1000$  zaczynające się i kończące zarówno w punkcie (1) oraz (2) tj. punktach, w których wykryto sygnał anty-korelacji.



Rysunek 3. Wykładnik Hursta wyliczony dla spółki Dębica wraz z wartością progową  $H_{threshold} = 0.4$ , oraz wartości kursów akcji i trendy liniowe.

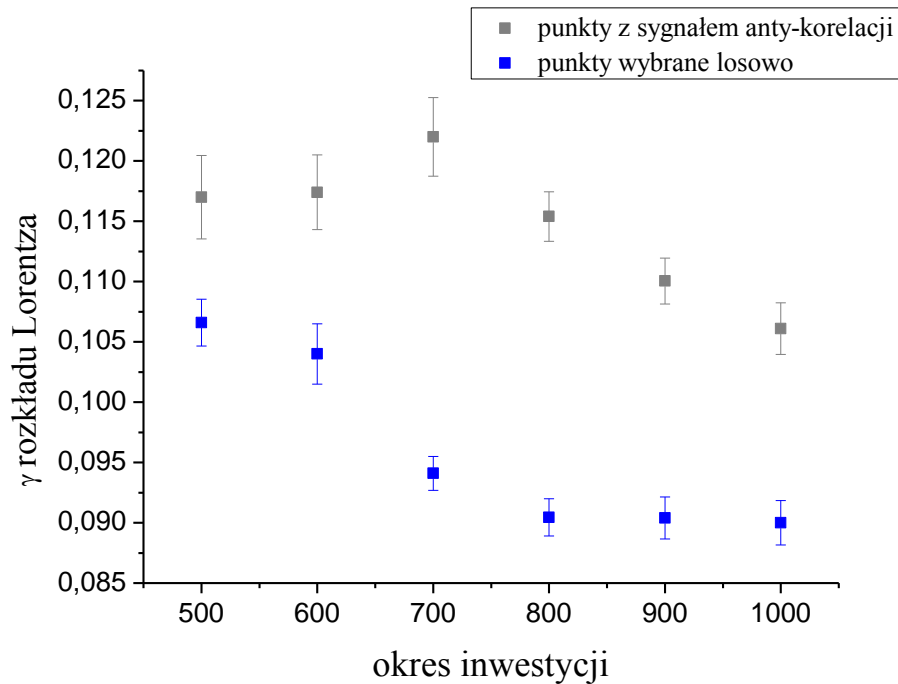
Jako istotną zmianę trendu przyjęto zmianę znaku jego nachylenia. Analizując rysunek 3, można stwierdzić, że w pkt. (1) nie nastąpiła zmiana trendu, natomiast w pkt. (2) nastąpiła. Dla badanych 126 spółek notowanych na GPW w Warszawie wyznaczono 240 punktów dla których  $H(t) < H_{threshold} = 0.4$ , zmianę znaku trendu liniowego stwierdzono w 51.3% przypadkach. Dla losowo wybranych punktów zmiana znaku trendu wystąpiła w 45.8% przypadkach. Okazało się ponadto, że

wraz ze spadkiem wielkości  $H_{threshold}$  udział punktów, w których wystąpiła zmiana znaku trendu (w stosunku do ogólnej ilości  $H(t) < H_{threshold}$ ) wzrastał, aż do 62% dla  $H_{threshold} = 0.275$ . Niestety odbywało się to przy znacznym spadku ilości zarejestrowanych sygnałów anty-korelacji [24]. Ogólnie wyciągnięto wniosek, że zmiana znaku trendu jest bardziej prawdopodobna w punktach, w których  $H < H_{threshold}$  niż w punktach wybranych losowo.

Następnie badano  $\Delta R_l(H < H_{threshold})$  - zmiany procentowych stóp zwrotu z inwestycji za okres  $l = 400, 500, \dots 1000$  w 240 punktach, w których wystąpił sygnał anty-korelacji, rozumiany jako  $H(t) < H_{threshold} = 0.4$ . Badano również  $\Delta R_l(\tau_{random})$  - procentowe zmiany stóp zwrotu w 240 punktach wybranych losowo. W obu przypadkach sporządzono rozkład gęstości prawdopodobieństwa oraz dopasowano rozkład Lorentza [24]. Rozkład Lorentza należy do rodziny rozkładów Levy'ego i był wykorzystywany do analizy danych finansowych [12]. Jedną z interesujących cech rozkładu Lorentza jest występowanie tzw. „grubych ogonów” - relatywnie dużych prawdopodobieństw wystąpienia wartości ekstremalnych. Występowanie zjawisk ekstremalnych, np. dużych spadków lub wzrostów stóp zwrotu, było obserwowane wielokrotnie dla danych finansowych [12]. Gęstość prawdopodobieństwa rozkładu Lorentza opisana jest równaniem:

$$\rho(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - x_0)^2}, \quad (37)$$

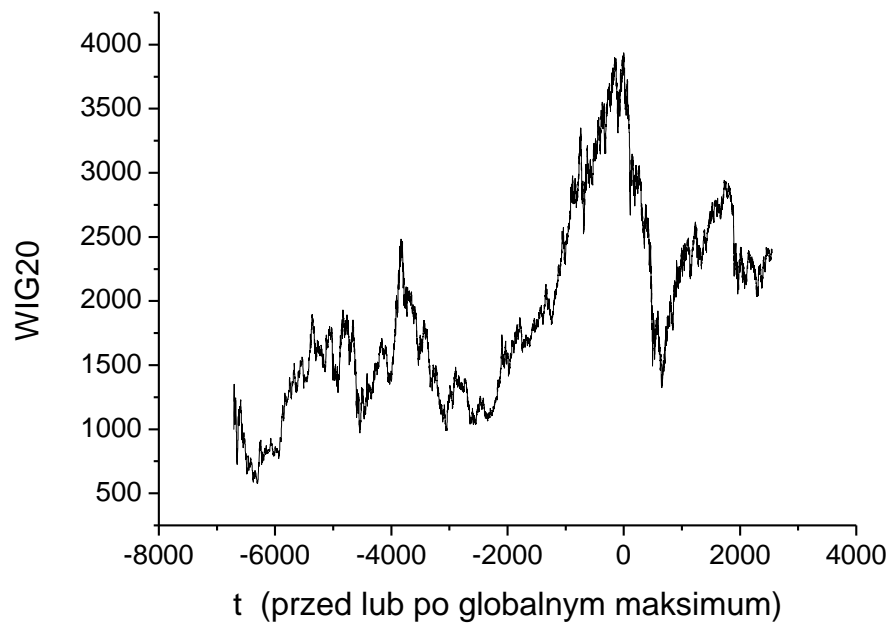
gdzie  $x_0$  to wartość modalna. Podczas badań stwierdzono, że  $x_0 \approx 0$ , co oznacza, że średnio statystycznie stopy zwrotu nie ulegają zmianie. Wartość współczynnika  $\gamma$  odpowiada za rozpiętość rozkładu prawdopodobieństwa; im większa jego wartość, tym występuje więcej wielkości ekstremalnych oraz mniej wielkości około - modalnych. Na rysunku 4 przedstawiono wartości  $\gamma$  otrzymane dzięki dopasowaniu krzywej Lorentza - słupki niepewności odpowiadają niepewności w wyznaczeniu wartości  $\gamma$ .



Rysunek 4. Wartości  $\gamma$  dla zmiany stóp zwrotu w punktach, w których wykryto sygnał anty-korelacji  $\Delta R_l(H < H_{threshold})$  oraz w punktach wybranych losowo -  $\Delta R_l(\tau_{random})$ .

Warto w tym miejscu stwierdzić, że wykładnik Hursta jest wskaźnikiem, który może okazać się pomocny w procedurze wyboru jednowymiarowego rozkładu prawdopodobieństwa dla danych finansowych, określającego np. ryzyko straty na jednym papierze wartościowym. Należy jednak pamiętać, że dla inwestorów znacznie bardziej istotne jest ryzyko straty na kilku papierach wartościowych należących do portfela inwestycyjnego [34, 35]. Do analizy takiego ryzyka tworzy się wielowymiarowe modele typu „VaR”, które zostaną omówione w późniejszej części pracy.

W kolejnej części dysertacji zostanie omówiona praca opublikowana przez autora [25], badająca wykorzystanie wykładnika Hursta do analizy sygnału anty-korelacji przed zbliżającym się maksimum indeksu WIG 20, które wystąpiło 29 października 2007 roku - rysunek 5.



Rysunek 5. Notowania indeksu WIG 20, założono  $t = 0$  podczas globalnego maksimum z 29 października 2007 roku.

Indeks WIG 20 odzwierciedla zachowanie się 20 najważniejszych spółek notowanych na GPW w Warszawie [33]. Wylicza się go na podstawie równania (38):

$$WIG20 = 1000 \frac{1}{K(t)} \frac{M(t)}{M(t=0)}, \quad (38)$$

gdzie  $K(t)$  to współczynnik korygujący, którego wartość zmienia się wraz ze zmianą składu indeksu, natomiast  $M(t)$  to kapitalizacja indeksu opisana równaniem (39):

$$M(t) = \sum_{i=1}^{20} P_i(t) S_i(t), \quad (39)$$

$P_i(t)$  i  $S_i(t)$  to cena i ilość wyemitowanych akcji  $i$ -tej spółki należącej do indeksu. Spółki biorące udział w indeksie WIG 20 ustala się na podstawie ostatniej sesji stycznia, kwietnia, lipca i października, jednakże czasem zdarzają się nadzwyczajne zmiany. W tabeli 2 przedstawiono udziały

spółek w indeksie WIG 20 z dnia 29 października 2007 roku, kiedy wystąpiło globalne maksimum. W dalszej części podrozdziału zostanie opisane badanie kursów akcji spółek ujętych w tabeli 2.

Tabela 2. Skład indeksu WIG 20 podczas globalnego maksimum z dnia 29 października 2007 r.

i	Spółka (i)	Udział w indeksie $w_i$	i	Spółka (i)	Udział w indeksie $w_i$
1	PKO BP	15%	11	POLIMEX MS	2.67%
2	PEKAO	13.89%	12	TVN	2.59%
3	PKN ORLEN	12.75%	13	PBG	2.2%
4	KGHM	10.98%	14	CEZ	2.11%
5	TPSA	9.09%	15	CERSANIT	2.06%
6	BANK BPH	5.26%	16	LOTOS	1.84%
7	BZ WBK	4.52%	17	BIOTON	1.55%
8	GTC	3.43%	18	AGORA	1.42%
9	BRE (MBANK)	3.35%	19	PROKOM	1.27%
10	PGNIG	3.29%	20	POLNORD	0.78%

Wielkości  $w_i$  to udziały spółek w indeksie WIG 20. Analizując tabelę 2 można zauważyć, że wartości te różnią się w znacznym stopniu dla różnych spółek. Przekształcając wzory (38, 39), przy założeniu stałości składu indeksu, która była zachowana w okolicach globalnego maksimum, otrzymujemy [25]:

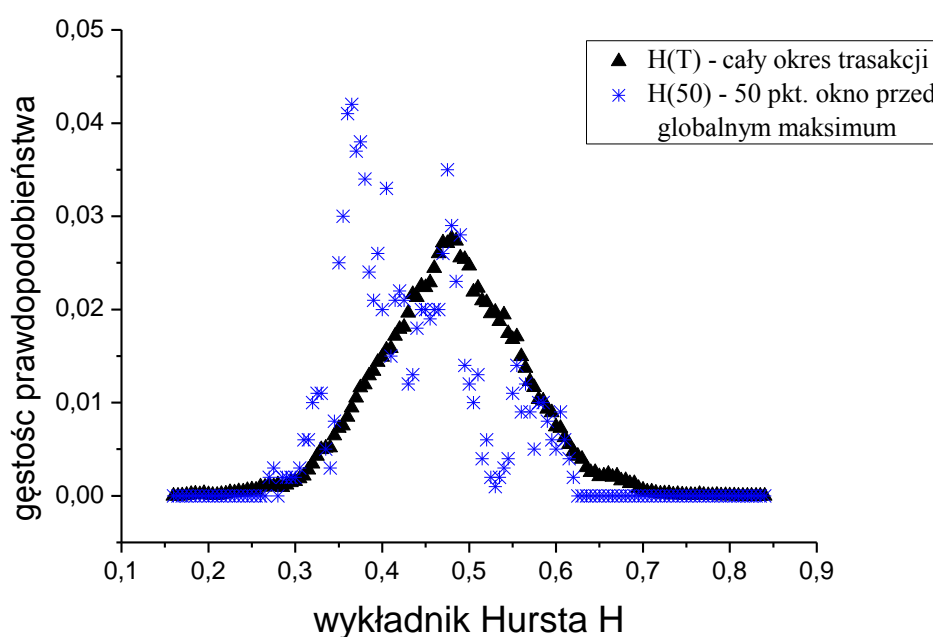
$$WIG20(t) = const \sum_{i=1}^{20} w_i P_i(t). \quad (40)$$

Indeks WIG 20 jest ważoną sumą kursów akcji, dlatego analizując wykładnik Hursta dla poszczególnych spółek należących do indeksu można badać sygnał anty-korelacji dla całego indeksu. Poniższą analizę przeprowadzono bez uwzględnienia udziałów spółek w indeksie  $w_i$ , analizując wartości wykładników Hursta wyliczone dla poszczególnych spółek. Jest to podejście przybliżone. Do analizy danych wykorzystano lokalny wykładnik Hursta  $H$  wyliczony z wykorzystaniem lokalnego DFA dla okien o długości  $T = 1000$ ,  $T = 500$ ,  $T = 250$ . Następnie zebrano razem wartości wykładników  $H$  wszystkich spółek tworząc zmienną jednowymiarową. Zmienną tę badano w oknach czasowych o rozmiarach od  $T' = 200$  do  $T' = 10$  przed oraz po globalnym maximum [25]. Z badań wynika, że w pobliżu globalnego maksimum wykładnik Hursta osiągał wartości  $H \ll 0.5$  sugerujący wystąpienie



sygnału anty-korelacji. Rozkład gęstości prawdopodobieństwa wykładnika  $H$ , dla wszystkich spółek należących do indeksu WIG 20 - patrz tabela 2 - przedstawiono na rysunku 6 dla następujących okien czasowych:

- całego okresu notowań spółek należących do indeksu WIG 20 -  $H_{250}(T)$ ,
- okna czasowego zaczynającego się 50 punktów (25 dni) przed globalnym maksimum oraz kończącym się w punkcie globalnego maksimum  $H_{250}(T' = 50)$ .



Rysunek 6. Gęstość prawdopodobieństwa dla  $H_{250}(T)$  oraz  $H_{250}(50)$  [25].

Analizując rysunek 6 można stwierdzić, że wartość modalna  $H_{250}(50)$  jest mniejsza niż wartość modalna  $H_{250}(T)$ , co sugeruje wystąpienie sygnału anty-korelacji przed globalnym maksimum. W pracy [25] zbadano ponadto wykładnik Hursta  $H^{shuff}$  dla danych powstałych po losowym wymieszaniu stóp zwrotu, aby sprawdzić, czy spadek wykładnika Hursta jest spowodowany sygnałem anty-korelacji, czy związany jest z innymi właściwościami (np. multifraktałnymi) szeregów czasowych, jak miało to miejsce w [36]. Badając rozkłady prawdopodobieństwa, okazało się że w okolicach globalnego maksimum indeksu WIG 20 zachodzi warunek  $H^{shuff}(t) > H(t)$  [25], co oznacza, że sygnał anty-korelacji jest silniejszy dla oryginalnych danych niż dla danych powstałych po przeprowadzeniu transformacji losującej. W pracy [25] przedstawiono również prosty sposób uwzględnienia udziałów poszczególnych spółek w indeksie  $w_i$ . Ocena ta z grubsza polegała na badaniu ważonych średnich wykładnika Hursta. Wykazała ona również występowanie sygnału

anty-korelacji w okolicach globalnego maksimum. Jednakże głębsza analiza wymagałaby stworzenia zmiennej wielowymiarowej modelującej zachowanie poszczególnych składników indeksu WIG 20.

Podjęcie przedstawione w poprzednim paragrafie zaniedbywało korelacje pomiędzy różnymi wykładnikami Hursta  $H(P_i)$  w funkcji  $H(P_j)$ , gdzie indeksy  $i$  oraz  $j$  numerują kolejne spółki należące do indeksu WIG 20 ( $i, j = 1, \dots, 20$ ). Ogólnie rzecz biorąc wykładnik  $H$  może być użyty jako jakościowy sygnał ostrzegawczy przed załamaniem się rynku finansowego. Wykorzystanie wykładnika w sposób ilościowy do oceny prawdopodobieństwa wystąpienia ujemnej zmiany cen akcji  $\Delta P(t)$  lub ujemnej stopy zwrotu wymaga dalszych badań. Badania takie dla jednej spółki lub jednego indeksu wymagają zastosowania odpowiednich modeli gęstości prawdopodobieństwa. W przypadku wielowymiarowym - kursów akcji wielu spółek - należy zastosować wielowymiarowy rozkład prawdopodobieństwa. Podczas tworzenia wielowymiarowych rozkładów prawdopodobieństwa bardzo przydatne okazują się funkcje Kopuły [1]. Funkcje te będą znajdować coraz większe zastosowanie w analizie układów finansowych, ponieważ analiza jednoczesnego prawdopodobieństwa spadku wielu kursów akcji jest w centrum zainteresowania osób zajmujących się zarządzaniem ryzykiem na rynkach finansowych. Dodatkowo, połączenie analizy DFA oraz funkcji Kopuły jest nowatorskim narzędziem i zakłada się, że będzie w przyszłości stosowane do analizy rynków finansowych oraz różnorodnych fizycznych układów złożonych. Funkcje Kopuły jako istotne w analizie danych finansowych zostaną omówione w następnym rozdziale.

Na koniec rozdziału warto wspomnieć o modelach mikroskopowych rynków finansowych. W literaturze można znaleźć wiele modeli mikroskopowych potwierdzających wystąpienie anomalnej dyfuzji dla szeregów czasowych złożonych z kursów akcji lub wartości indeksów giełdowych [2]. Jako pierwszy przykład można podać model Bak-Paczuski [37], który zakłada, iż za duże wahania kursów akcji odpowiada efekt tłumu, gdy jedni inwestorzy naśladują zachowanie innych. Model bada różne zachowania się rynku finansowego. Jeżeli przewagę mają gracze „racjonalni” kierujący się analizą fundamentalną oraz realną wartością spółek, zmienność kursów akcji jest relatywnie niewielka. Jeżeli przewagę mają gracze „tworzący szum”, którzy kierują się „dynamiką rynku” oraz naśladują zachowanie innych graczy tworzą się „bańki” spekulacyjne, a wahania kursów akcji są znacznie większe niż te przewidziane przez analizę fundamentalną. Następnie warto wspomnieć o modelu inwestorów znajdujących się pod „presją” rynku [38], który również dzieli inwestorów na dwie grupy - inwestorów racjonalnych oraz inwestorów nieracjonalnych „tworzących szum”. Między inwestorami tworzą się napięcia oraz „tarcia” związane z różną strategią oraz horyzontem inwestycyjnym. Z modelu można wywnioskować, że pojawienie się inwestorów nieracjonalnych powoduje „subdyfuzyjne” zachowanie się rynku -  $H < 0.5$ . Na koniec warto wspomnieć o wykorzystaniu modelu Isinga do modelowania załamania się rynku finansowego w analogii to przejścia fazowego [39]. W modelu tym zakłada się, że poszczególni inwestorzy są zorganizowani na podstawie sieci

(rodzina, koledzy, przyjaciele, itp.) i każdy z nich ma ustaloną ilość najbliższych sąsiadów z którymi „oddziałuje”. Inwestorzy mogą sugerować się zachowaniem innych lub stosować strategie indywidualne. Model ten dostarcza ciekawych wyników sugerujących pojawianie się „bąbli” spekulacyjnych oraz załamania rynku w analogii do przejścia fazowego pierwszego rodzaju.

### 3.4. Funkcje Kopuły

#### 3.4.1 Definicja, zastosowanie i historia

W rozdziale tym zostaną omówione funkcje Kopuły, które znajdują szerokie zastosowanie w badaniu wielowymiarowych zmiennych losowych. Funkcje Kopuły zostały po raz pierwszy użyte w 1959 roku [40] w celu badania relacji pomiędzy wielowymiarowym rozkładem prawdopodobieństwa oraz jednowymiarowymi (marginalnymi) rozkładami prawdopodobieństwa. Funkcje te z biegiem czasu znalazły zastosowanie w analizie słabo przewidywanych danych, w tym między innymi w:

- kosmologii, badaniu gęstości materii w wielkiej skali – np. badaniu rozkładu gęstości ciemnej materii [41],
- badaniu pogody, np. badaniu zależności pomiędzy temperaturą i opadami [42, 43],
- medycynie - tomografii, badaniu kory mózgowej [44, 45],
- informatyce - generowaniu wektorów losowych [46],
- finansach - ekonometrii, ocenie i zarządzaniu ryzykiem, badaniu ryzyka kredytowego [1].

Funkcje Kopuły ( $C$ ) są zdefiniowane na  $n$  wymiarowym jednostkowym odcinku  $[0,1]^n$  poprzez transformację na jednowymiarowy odcinek  $[0,1]$ , co można zobrazować przez przekształcenie  $C: [0,1]^n \rightarrow [0,1]$ . Funkcje te stosuje się w konstruowaniu wielowymiarowego rozkładu prawdopodobieństwa. Rozważmy wielowymiarową zmienną losową  $[x_1, \dots, x_i, \dots, x_n]$ , każdy element zmiennej losowej -  $x_i$  posiada jednowymiarowy „brzegowy” rozkład prawdopodobieństwa  $f_i(x_i)$  oraz dystrybuantę  $F_i(x_i)$ . Zdefiniujmy wielowymiarową zmienną  $[F_1(x_1), \dots, F_i(x_i), \dots, F_n(x_n)] = [v_1, \dots, v_i, \dots, v_n]$  która dzięki funkcji Kopuły może zostać przetransformowana do przestrzeni  $[0,1]$  tworząc dystrybuantę wielowymiarowego rozkładu prawdopodobieństwa:

$$C: [F_1(x_1), \dots, F_i(x_i), \dots, F_n(x_n)] \rightarrow [0,1]. \quad (41)$$

Procedurę dopasowania funkcji Kopuły omówiono szczegółowo w podrozdziale 3.4.4. W praktyce procedura utworzenia wielowymiarowej dystrybuanty przy pomocy funkcji Kopuły polega na:

- dopasowaniu jednowymiarowych rozkładów brzegowych  $f_i(x_i)$ ,
- wyznaczeniu brzegowych dystrybuant  $F_i(x_i)$ ,
- zmianie zmiennych  $v_i = F_i(x_i)$ ,
- wyborze odpowiedniej funkcji Kopuły  $C$ ,
- dopasowaniu parametrów funkcji Kopuły  $\theta_1, \dots, \theta_m$  do danych  $C(\theta_1, \dots, \theta_m): [v_1, \dots, v_i, \dots, v_n]$ .

Funkcje Kopuły muszą spełnić kilka istotnych warunków [1]:

- $C[F_1(x_1), \dots, F_i(x_i) = 0, \dots, F_n(x_n)] = 0$  - jeżeli jedna dystrybuanta brzegowa jest równa zero, prawdopodobieństwo wystąpienia  $i$ -tej wartości mniejszej lub równej  $x_i$  jest również równe zero, dlatego prawdopodobieństwo wystąpienia zmiennej wielowymiarowej mniejszej lub równej  $[x_1, \dots, x_i, \dots, x_n]$  również musi być równe zero;
- $C[F_1(x_1) = 1, \dots, 0 < F_i(x_i) < 1, \dots, F_n(x_n) = 1] = F_i(x_i)$  - jeżeli prawdopodobieństwo wystąpienia  $n - 1$  wymiarowej zmiennej losowej (bez zmiennej na  $i$ -tej pozycji) mniejszej bądź równej  $[x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$  jest równe jeden to prawdopodobieństwo wystąpienia  $n$  wymiarowej zmiennej losowej (ze zmienną na  $i$ -tej pozycji) mniejszej bądź równej  $[x_1, \dots, x_i, \dots, x_n]$  jest równe prawdopodobieństwu wystąpienia zmiennej na  $i$ -tej pozycji mniejszej lub równej  $x_i$ ;
- $C[F_1(x_1), \dots, F_i^1(x_i), \dots, F_n(x_n)] < C[F_1(x_1), \dots, F_i^2(x_i), \dots, F_n(x_n)]$  jeżeli  $F_i^1(x_i) < F_i^2(x_i)$  - funkcja Kopuły musi być  $n$ -rosnąca - dla każdego  $i$  wzrost wartości argumentu na  $i$ -tej pozycji powoduje wzrost wartości funkcji Kopuły - jeżeli wzrośnie wartość dystrybuanty marginalnej musi również wzrosnąć wartość dystrybuanty całosciowej obliczonej z użyciem funkcji Kopuły;
- podsumowując funkcja Kopuły działa na odcinek  $[0,1]^n$  dając w wyniku odcinek  $[0,1]$  w ten sposób, że  $C: [1,1, \dots, 1] = 1$  oraz  $C: [0,0, \dots, 0] = 0$ .

### 3.4.2 Podział i rodzaje funkcji Kopuły

Jedną z podstawowych funkcji Kopuły jest Kopuła Gaussowska [1]. Aby zdefiniować Kopułę Gaussowską, założymy, że  $\Phi_R(v_1, \dots, v_i, \dots, v_n)$  to zestandaryzowana, wielowymiarowa dystrybuanta Gaussa z macierzą korelacji  $R$  równą:

$$R = \begin{bmatrix} 1 & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{1n} & \cdots & 1 \end{bmatrix} \quad (42)$$

gdzie  $\rho_{ij}$  to współczynnik korelacji pomiędzy  $i$ -tą oraz  $j$ -tą zmienną. Załóżmy ponadto, że  $\Phi^{-1}(v)$  to odwrotna dystrybuanta zestandaryzowanego jednowymiarowego rozkładu Gaussa. Kopułę Gaussowską możemy zapisać w postaci:

$$C^{Ga}(v_1, \dots, v_i, \dots, v_n) = \Phi_R(\Phi^{-1}(v_1), \dots, \Phi^{-1}(v_i), \dots, \Phi^{-1}(v_n)). \quad (43)$$

W dwóch wymiarach mamy:

$$R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (44)$$

oraz

$$C^{Ga}(v_1, v_2) = \Phi_R(\Phi^{-1}(v_1), \Phi^{-1}(v_2)). \quad (45)$$

W tym miejscu nasuwa się pytanie, jaka jest różnica pomiędzy Kopułą Gaussowską -  $C^{Ga}$  a wielowymiarowym rozkładem Gaussa -  $\Phi_R$ . Wyobraźmy sobie wielowymiarową zmienną losową  $[x_1, \dots, x_i, \dots, x_n]$ , dla której każdy składnik posiada brzegowy rozkład prawdopodobieństwa będący zestandaryzowanym rozkładem Gaussa -  $F_i(x_i) = \Phi(x_i)$ . Transformujemy teraz tę zmienną do przestrzeni  $[0,1]^n$ :

$$v_i = F_i(x_i) = \Phi(x_i), \quad (46)$$

i podstawiamy do wzoru na Kopułę Gaussowską,

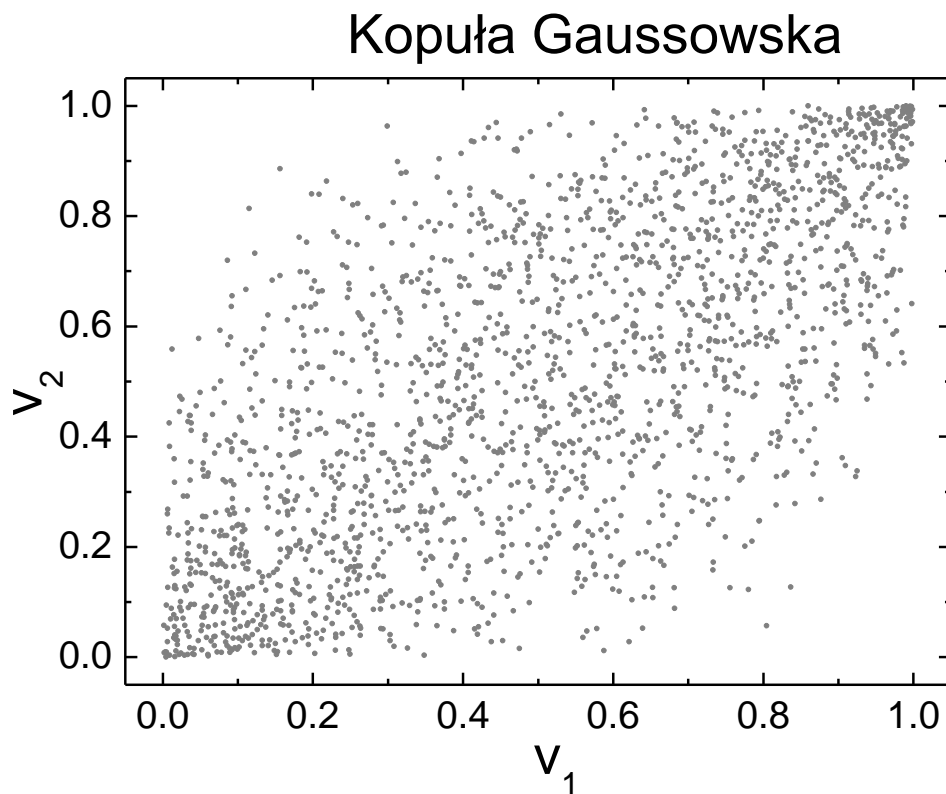
$$C^{Ga}(v_1, \dots, v_i, \dots, v_n) = \Phi_R(\Phi^{-1}(\Phi(x_1)), \dots, \Phi^{-1}(\Phi(x_i)), \dots, \Phi^{-1}(\Phi(x_n))) = \Phi_R(x_1, \dots, x_i, \dots, x_n). \quad (47)$$

Oznacza to, że w tym przypadku Kopuła Gaussowska prowadzi do wielowymiarowego rozkładu Gaussa. Jeśli jednak rozkłady brzegowe  $f_i(x_i)$  nie są rozkładami Gaussa, również dystrybuanty brzegowe nie będą dystrybuantami Gaussa -  $F_i(x_i) \neq \Phi(x_i)$ . Sytuacja taka może mieć miejsce np.

podczas badania zmiennej ekstremalnej opisanej przez Rozkład Wartości Ekstremalnej różniący się znacząco od rozkładu Gaussa (rozdział 4.1). Podobne rozumowanie można przeprowadzić, jeżeli nie wszystkie rozkłady marginalne dla  $n$ -wymiarowej zmiennej są rozkładami Gaussa - dla różnych  $i$  mogą występować różne typy rozkładów brzegowych  $f_i(x_i)$ . Ogólnie, jeżeli zachodzi przynajmniej jedna z relacji  $F_i(x_i) \neq \Phi(x_i)$ , wtedy równanie (47) przyjmuje postać:

$$C^{Ga}(v_1, \dots, v_i, \dots, v_n) = \Phi_R \left( \Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_i(x_i)), \dots, \Phi^{-1}(F_n(x_n)) \right) \neq \Phi_R(x_1, \dots, x_i, \dots, x_n). \quad (48)$$

W tym przypadku Kopuła Gaussowska nie odpowiada wielowymiarowemu rozkładowi Gaussa. Reasumując, Kopuła Gaussowska jest uogólnieniem wielowymiarowego rozkładu Gaussa pozwalającym na zastosowanie rozkładów brzegowych innych niż rozkłady Gaussa. Na rysunku 7 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez dwuwymiarową Kopułę Gaussowską ze współczynnikiem korelacji  $\rho = 0.7$ .



Rysunek 7. Dane wygenerowane przez dwuwymiarową Kopułę Gaussowską.

Podobnie można zdefiniować wielowymiarową Kopułę Studenta [1, 47]. Załóżmy, że  $\mathbf{t}_{R\nu}$  to znormalizowana wielowymiarowa dystrybuanta Studenta:

$$\mathbf{t}_{R\nu}(x_1, \dots, x_i, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_i} \dots \int_{-\infty}^{x_n} \frac{\Gamma\left(\frac{\nu+n}{2}\right) |R|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{\frac{n}{2}}} \left(1 + \frac{1}{\nu} \mathbf{x}^T R^{-1} \mathbf{x}\right)^{-\frac{\nu+n}{2}} dx_1 \dots dx_i \dots dx_n, \quad (49)$$

gdzie  $R$  to macierz korelacji, natomiast  $\nu$  to ilość stopni swobody. Zmienna wielowymiarowa zapisana jest w postaci wektora  $\mathbf{x} = [x_1, \dots, x_i, \dots, x_n]$ . Kopuła Studenta jest zdefiniowana w następujący sposób:

$$C^T(v_1, \dots, v_i, \dots, v_n) = \mathbf{t}_{R\nu}(t_v^{-1}(v_1), \dots, t_v^{-1}(v_i), \dots, t_v^{-1}(v_n)) \quad , \quad (50)$$

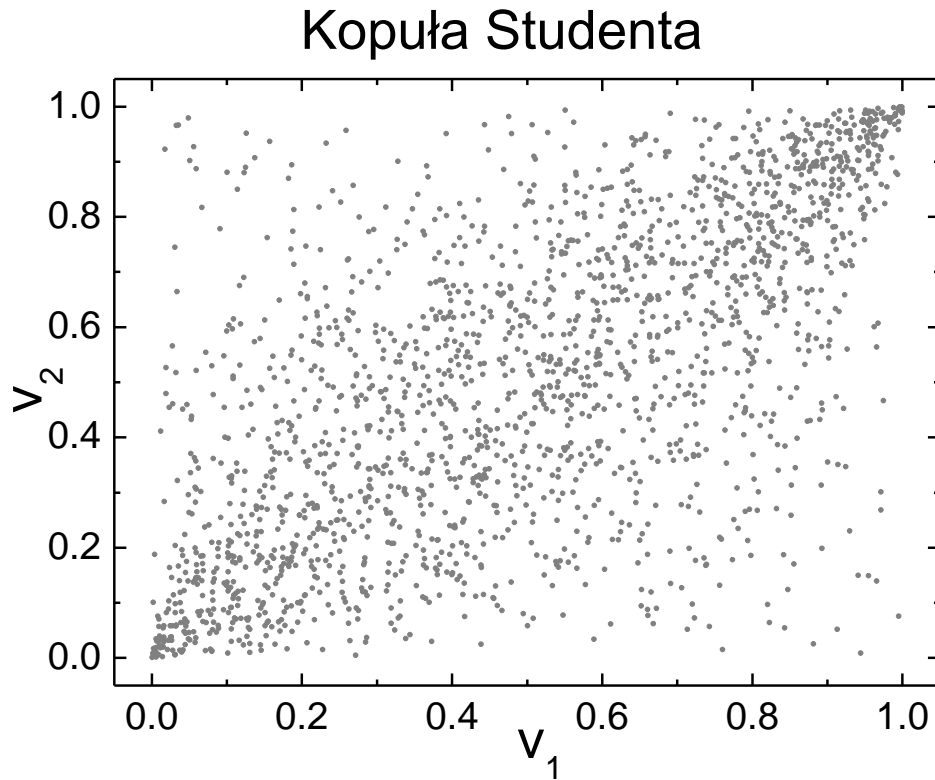
gdzie  $t_v^{-1}$  to odwrotna dystrybuanta zestandaryzowanego jednowymiarowego rozkładu Studenta. W przypadku dwuwymiarowym mamy jeden współczynnik korelacji  $\rho$  ujęty w macierzy korelacji - wzór (44) – a Kopuła Studenta  $C_{\rho,\nu}^T(v_1, v_2)$  zależy od dwóch parametrów: wspomnianego współczynnika korelacji  $\rho$  oraz ilości stopni swobody. Tak jak w przypadku Kopuły Gaussowskiej, Kopuła Studenta [1, 47] prowadzi do wielowymiarowego rozkładu Studenta, jeżeli wszystkie brzegowe rozkłady są rozkładami Studenta posiadającymi taką samą ilość stopni swobody jak badana Kopuła Studenta, czyli  $F_i(x_i) = t_\nu(x_i)$ :

$$C^T(v_1, \dots, v_i, \dots, v_n) = \mathbf{t}_{R\nu}(t_v^{-1}(t_\nu(x_1)), \dots, t_v^{-1}(t_\nu(x_i)), \dots, t_v^{-1}(t_\nu(x_n))) = \mathbf{t}_{R\nu}(x_1, \dots, x_i, \dots, x_n). \quad (51)$$

W przeciwnym razie, jeśli  $F_i(x_i) \neq t_\nu(x_i)$ , Kopuła Studenta nie prowadzi do wielowymiarowego rozkładu Studenta. Zatem:

$$C^T(v_1, \dots, v_i, \dots, v_n) = \mathbf{t}_{R\nu}(t_v^{-1}(F_1(x_1)), \dots, t_v^{-1}(F_i(x_i)), \dots, t_v^{-1}(F_n(x_n))) \neq \mathbf{t}_{R\nu}(x_1, \dots, x_i, \dots, x_n). \quad (52)$$

Na rysunku 8 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez dwuwymiarową Kopułę Studenta ze współczynnikiem korelacji  $\rho = 0.7$ , oraz  $\nu = 3$ .



Rysunek 8. Dane wygenerowane przez dwuwymiarową Kopułę Studenta

W tym miejscu warto wspomnieć o rodzinie Kopuł Archimedesowskich -  $C^A(v_1, \dots, v_i, \dots, v_n)$ . Kopuły te są zdefiniowane przez funkcję  $\phi$  będącą ich generatorem [1], która spełnia następujące warunki:

- $\phi(1) = 0$ ,
- funkcja  $\phi$  jest ciągła, malejąca i wypukła,
- jeżeli ponadto  $\phi(0) = \infty$  mamy odczynienia z pełnym generatorem.

Jeśli powyższe warunki są spełnione funkcja opisana przez poniższy wzór jest funkcją Kopuły, to jest

$$C^A(v_1, \dots, v_i, \dots, v_n) = \phi^{[-1]}(\phi(v_1) + \dots + \phi(v_i) + \dots + \phi(v_n)), \quad (53)$$

gdzie  $\phi^{[-1]}$  to pseudo-funkcja odwrotna [1] opisana wzorem:

$$\phi^{[-1]}(x) = \begin{cases} \phi^{-1}(x) & 0 \leq x \leq \phi(0) \\ 0 & \phi(0) \leq x \leq +\infty. \end{cases} \quad (54)$$



Jak widać w przypadku, kiedy funkcja  $\phi$  jest pełnym generatorem możemy zapisać:

$$C^A(v_1, \dots, v_i, \dots, v_n) = \phi^{-1}(\phi(v_1) + \dots + \phi(v_i) + \dots + \phi(v_n)). \quad (55)$$

Poniżej zostaną przedstawione najbardziej istotne Kopuły Archimedesowskie.

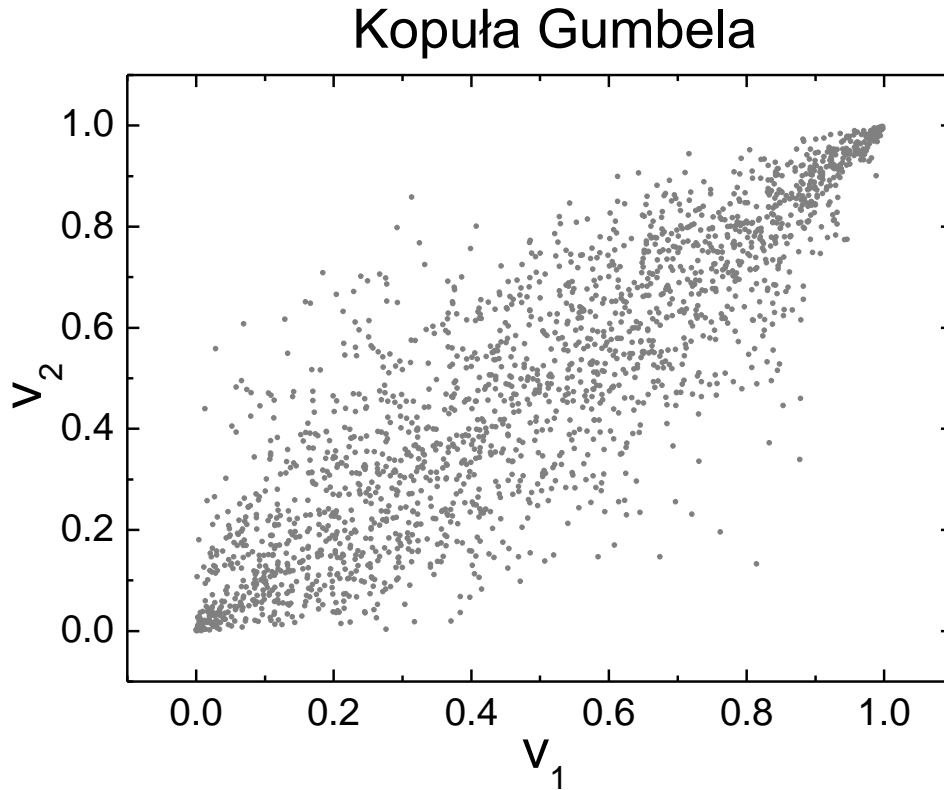
Rozważmy w pierwszej kolejności **Kopułę Gumbela**, posiadającą generator  $\phi(v) = (-\ln v)^\theta$ , oraz funkcję odwrotną generatora  $\phi^{-1}(x) = e^{-x^{\frac{1}{\theta}}}$ . Dla  $\theta \geq 1$  funkcja  $\phi(v)$  jest pełnym generatorem. Kopuła Gumbela jest określona dla  $\theta > 1$  w przypadku wielowymiarowym [1] jako:

$$C^G(v_1, \dots, v_i, \dots, v_n) = \exp\left(-\left(\sum_{i=1}^n (-\ln v_i)^\theta\right)^{\frac{1}{\theta}}\right). \quad (56)$$

W przypadku dwuwymiarowym Kopuła ta jest określona dla  $\theta \geq 1$  [1] poprzez:

$$C_\theta^G(v_1, v_2) = \exp\left(-((- \ln v_1)^\theta + (- \ln v_2)^\theta)^{\frac{1}{\theta}}\right). \quad (57)$$

Na rysunku 9 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez dwuwymiarową Kopułę Gumbela z parametrem  $\theta = 3.33$ .



Rysunek 9. Dane wygenerowane przez dwuwymiarową Kopułę Gumbela.

Rozważmy **Kopułę Claytona** posiadającą następujący generator  $\phi(v) = \frac{1}{\theta}(v^{-\theta} - 1)$  oraz jego funkcję odwrotną  $\phi^{-1}(x) = (1 + \theta x)^{-\frac{1}{\theta}}$ . W przypadku wielowymiarowym ( $d \geq 3$ ) Kopuła Claytona jest  $n$ -rosnąca dla  $\theta \geq -\frac{1}{d-1}$  [48], co jest zawsze spełnione dla  $\theta > 0$ . Ponadto w przypadku  $\theta > 0$  funkcja  $\phi^{-1}$  jest pełnym generatorem a Kopuła Claytona przyjmuje postać [1]:

$$C^C(v_1, \dots, v_i, \dots, v_n) = \left( \sum_{i=1}^n v_i^{-\theta} - n + 1 \right)^{-\frac{1}{\theta}}. \quad (58)$$

W przypadku dwuwymiarowym ( $d = 2$ ) Kopuła Claytona jest monotoniczna dla  $\theta \geq -1 \wedge \theta \neq 0$ . Jeżeli rozważymy generator  $\phi(v) = \frac{1}{\theta}(v^{-\theta} - 1)$ , posiadający funkcję odwrotną  $\phi^{-1} = (1 + \theta x)^{-\frac{1}{\theta}}$  dla  $\theta < 0$ , wtedy otrzymujemy:

$$\phi(0) = \frac{1}{\theta}(0^{-\theta} - 1) = -\frac{1}{\theta}, \quad (59)$$

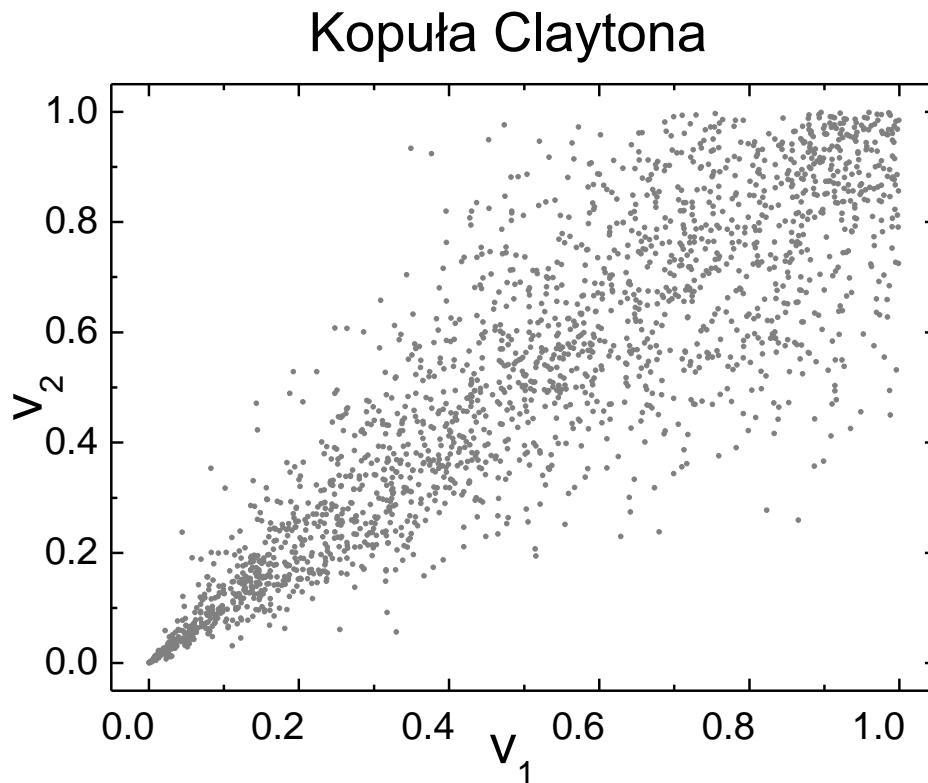
co jest liczbą skończoną. Dlatego należy użyć pseudo-wartości odwrotnej  $\phi^{[-1]}(x)$  zgodnie z równaniem:

$$\phi^{[-1]}(x) = \begin{cases} \phi^{-1}(x) & 0 \leq x \leq \phi(0) \\ 0 & \phi(0) \leq x \leq +\infty \end{cases} \quad (60)$$

co daje Kopułę Claytona równą:

$$C_{\theta}^C(v_1, v_2) = \max\left(\left(v_1^{-\theta} + v_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}, 0\right). \quad (61)$$

Na rysunku 10 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez dwuwymiarową Kopułę Claytona z parametrem  $\theta = 4.47$ .



Rysunek 10. Dane wygenerowane przez dwuwymiarową Kopułę Claytona.

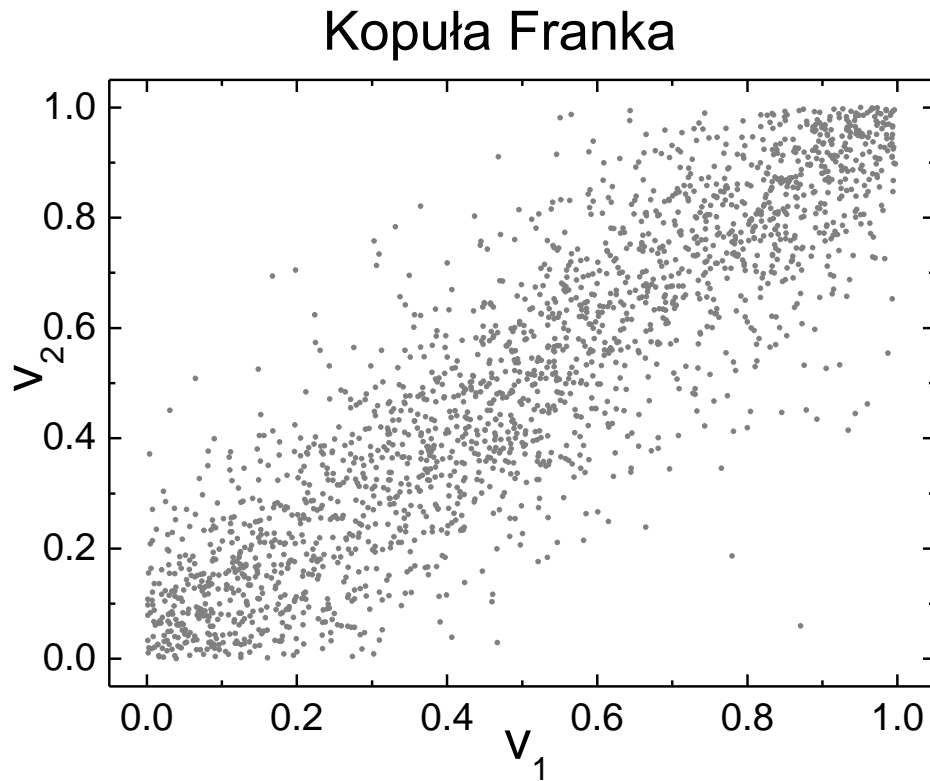
W tym miejscu można wspomnieć o **Kopule Franka**, rozważając następujący generator  $\phi(v) = \ln\left(\frac{\exp(-\theta v)-1}{\exp(-\theta)-1}\right)$  oraz jego odwrotność  $\phi^{-1}(x) = -\frac{1}{\theta} \ln\left(1 + e^x(e^{-\theta} - 1)\right)$ . W przypadku wielowymiarowym ( $d \geq 3$ ) Kopuła Franka jest określona dla  $\theta > 0$  poprzez [1]:

$$C^F(v_1, \dots, v_i, \dots, v_n) = -\frac{1}{\theta} \ln\left(1 + \frac{\prod_{i=1}^n (e^{-\theta v_i} - 1)}{(e^{-\theta} - 1)^{n-1}}\right). \quad (62)$$

W przypadku dwuwymiarowym Kopuła Franka jest określona dla  $\theta \in (-\infty, 0) \cup (0, \infty)$  poprzez [1]:

$$C_{\theta}^F(v_1, v_2) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta v_1} - 1)(e^{-\theta v_2} - 1)}{e^{-\theta} - 1}\right). \quad (63)$$

Na rysunku 11 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez dwuwymiarową Kopułę Franka z parametrem  $\theta = 11.41$ .



Rysunek 11. Dane wygenerowane przez dwuwymiarową Kopułę Franka.

Warto jeszcze wspomnieć o gęstości Kopuły zdefiniowanej przez wzór (64) [1]:

$$c(v_1, \dots, v_i, \dots, v_n) = \frac{\partial^n C(v_1, \dots, v_i, \dots, v_n)}{\partial v_1 \dots \partial v_i \dots \partial v_n}, \quad (64)$$

która jest związana z wielowymiarową gęstością prawdopodobieństwa  $f(x_1, \dots, x_i, \dots, x_n)$  poprzez reprezentację kanoniczną [1]:

$$f(x_1, \dots, x_i, \dots, x_n) = c(F_1(x_1), \dots, F_i(x_i), \dots, F_n(x_n)) \prod_{j=1}^n f_j(x_j), \quad (65)$$

gdzie  $f_j(x_j) = \frac{\partial F_j(x_j)}{\partial x_j}$  to gęstości rozkładów brzegowych. W przypadku dwuwymiarowym gęstość Kopuły [1] przyjmuje postać:

$$c(v_1, v_2) = \frac{\partial^2 C(v_1, v_2)}{\partial v_1 \partial v_2}, \quad (66)$$

natomiast jej reprezentacja kanoniczna:

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))(f_1(x_1)f_2(x_2)). \quad (67)$$

Następnie zostaną omówione właściwości wyżej wymienionych funkcji Kopuły w przypadku dwuwymiarowym. Funkcje te zostaną potem użyte do analizy danych.

### 3.4.3 Właściwości Kopuł dwuwymiarowych

W podrozdziale tym omówiono właściwości Kopuł dwuwymiarowych, które będą wykorzystane do dalszej analizy danych. Zwróćmy uwagę, na samym początku, na tzw. współczynnik korelacji rang  $\tau$  Kendall'a [1] zdefiniowany poprzez:

$$\tau(X_1, X_2) = E \left( \text{sign} \left( (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) \right) \right), \quad (68)$$

gdzie zarówno  $(X_1, X_2)$  jak i  $(\widetilde{X}_1, \widetilde{X}_2)$  to losowe wektory z badanego dwuwymiarowego rozkładu, natomiast  $E$  to wartość przewidywana. Ponieważ dystrybuanty brzegowe  $F_i(x_i)$  są funkcjami monotonicznie rosnącymi, wartość  $\tau$  Kendall'a będzie taka sama dla danych pierwotnych  $[x_1, x_2]$  jak dla danych przetransformowanych przy pomocy dystrybuant rozkładów brzegowych  $[v_1 = F_1(x_1), v_2 = F_2(x_2)]$ , co można zapisać następująco:

$$\tau(X_1, X_2) = \tau(V_1, V_2). \quad (69)$$

Współczynnik korelacji  $\tau$  Kendall'a można powiązać przy pomocy wzoru (70) z funkcją Kopuły [1]:

$$\tau = 4 \iint_{I^2} C(v_1, v_2) dC(v_2, v_2) - 1, \quad (70)$$

gdzie całkowanie odbywa się po jednostkowym kwadracie  $I^2 = [0,1]^2$ . Dla dwuwymiarowych Kopuł Archimedesowskich współczynnik  $\tau$  Kendall'a można bezpośrednio powiązać z generatorem Kopuły [1], według wzoru (71):

$$\tau = \int_I \frac{\phi(v)}{\phi'(v)} dv + 1. \quad (71)$$

Rozwiązując wzór (71) można wyznaczyć zależność pomiędzy współczynnikiem korelacji  $\tau$  Kendall'a a parametrem Kopuły  $\theta$  dla poszczególnych Kopuł Archimedesowskich [1]:

$$\begin{cases} \tau = 1 - \theta^{-1} & \text{Kopuła Gumbela} \\ \tau = \frac{\theta}{\theta+2} & \text{Kopuła Claytona} \\ \tau = 1 + \frac{4(D_1(\theta)-1)}{\theta} & \text{Kopuła Franka,} \end{cases} \quad (72)$$

gdzie  $D_1$  to odpowiednia funkcja „Debye'a”. Relacje te pozwalają na wyznaczenie parametru odpowiedniej Kopuły, jeśli wyznaczymy wcześniej współczynnik korelacji  $\tau$  Kendall'a.

Istotną właściwością funkcji Kopuły jest zależność ogonowa, która została wyprowadzona w celu badania jednoczesnego występowania zjawisk ekstremalnych. Jako indeks górnej zależności ogonowej ( $\lambda_u$ ) dwuwymiarowej Kopuły  $C(v_1, v_2)$  rozumiemy warunkowe prawdopodobieństwo wystąpienia dużej wartości pierwszej zmiennej ( $v_1 \rightarrow 1$ ), pod warunkiem wystąpienia dużej wartości drugiej zmiennej ( $v_2 \rightarrow 1$ ), co dla wartości  $\alpha \rightarrow 1$  można zapisać następująco:

$$\lambda_u = \Pr(v_1 \geq \alpha | v_2 \geq \alpha) = \frac{\Pr(v_1 \geq \alpha, v_2 \geq \alpha)}{\Pr(v_2 \geq \alpha)}. \quad (73)$$

Indeks  $\lambda_u$  można wyznaczyć znając równanie funkcji Kopuły [1], co pokazano we wzorze (74):

$$\lambda_u = \lim_{v \rightarrow 1^-} \frac{1 - 2v + C(v, v)}{1 - v}. \quad (74)$$

W tabeli 3 podano wartości indeksów dla omawianych Kopuły [1, 47], gdzie  $t_{v+1}$  to jednowymiarowa dystrybuanta studenta z  $v + 1$  stopniami swobody.

Tabela 3. Wartości indeksów górnej zależności ogonowej dla omawianych funkcji Kopuły.

Kopuła	$\lambda_u$
Gausa $C_\rho^{Ga}(v_1, v_2)$	0
Studenta $C_{\rho, v}^T(v_1, v_2)$	$2t_{v+1} \left( -\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}} \right)$
Gumbela $C_\theta^G(v_1, v_2)$	$2 - 2^{\frac{1}{\theta}}$
Claytona $C_\theta^C(v_1, v_2)$	0
Franka $C_\theta^F(v_1, v_2)$	0

Ponieważ zarówno  $v_1$  jak  $v_2$  mają rozkład jednorodny na odcinku  $[0,1]$ , pełniąc rolę dystrybuant rozkładów brzegowych, aby ułatwić analizę danych empirycznych zaproponowano przekształcenie wzoru (73) do postaci (75) przy wartości  $\alpha$  bliskiej 1,

$$\lambda_u = \Pr(v_1 \geq \alpha | v_2 \geq \alpha) = \frac{\Pr(v_1 \geq \alpha, v_2 \geq \alpha)}{1 - \alpha}. \quad (75)$$

Podobnie możemy zdefiniować indeks dolnej zależności ogonowej ( $\lambda_l$ ) dwuwymiarowej Kopuły  $C(v_1, v_2)$ , jako warunkowe prawdopodobieństwo wystąpienia małej wartości pierwszej zmiennej ( $v_1 \rightarrow 0$ ), pod warunkiem wystąpienia małej wartości drugiej zmiennej ( $v_2 \rightarrow 0$ ). Można to zapisać dla  $\alpha \rightarrow 0$  poprzez wzór (76):

$$\lambda_l = \Pr(v_1 \leq \alpha | v_2 \leq \alpha) = \frac{\Pr(v_1 \leq \alpha, v_2 \leq \alpha)}{\Pr(v_2 \leq \alpha)}, \quad (76)$$

gdzie zależność indeksu od funkcji Kopuły [1] przedstawia się następująco:

$$\lambda_l = \lim_{v \rightarrow 0^+} \frac{C(v, v)}{v}. \quad (77)$$

W tabeli 4 pokazano wartości indeksu dla omawianych Kopuł [1, 47].

Tabela 4. Wartości indeksów dolnej zależności ogonowej dla omawianych funkcji Kopuły.

Kopuła	$\lambda_l$
Gaussa $C_\rho^{Ga}(v_1, v_2)$	0
Studenta $C_{\rho, \nu}^T(v_1, v_2)$	$2t_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right)$
Gumbela $C_\theta^G(v_1, v_2)$	0
Claytona $C_\theta^C(v_1, v_2)$	$2^{-\frac{1}{\theta}}$
Franka $C_\theta^F(v_1, v_2)$	0

Podobnie, aby ułatwić analizę danych empirycznych zaproponowano przekształcenie wzoru (76) do następującej postaci, przy wartości  $\alpha$  bliskiej 0, mianowicie:

$$\lambda_l = \Pr(v_1 \leq \alpha | v_2 \leq \alpha) = \frac{\Pr(v_1 \leq \alpha, v_2 \leq \alpha)}{\alpha}. \quad (78)$$

Aby można było zastosować Kopułę Claytona do badania danych posiadających górną zależność ogonową lub Kopułę Gumbela do badania danych posiadających dolną zależność ogonową, zaproponowano wprowadzenie **odwróconej Kopuły Claytona** -  $C_\theta^{RC}(v_1, v_2)$  oraz **odwróconej Kopuły Gumbela** -  $C_\theta^{RG}(v_1, v_2)$ . Kopuły te zostały wyprowadzone z Kopuły Gumbela - wzór (57) - lub Kopuły Claytona - wzór (61) - poprzez transformację:

$$\begin{cases} v_1 \rightarrow 1 - v_1, \\ v_2 \rightarrow 1 - v_2. \end{cases} \quad (79)$$

Przypomnijmy, że zarówno  $v_1$  jak i  $v_2$  są zmiennymi jednorodnymi w przestrzeni  $[0,1]$ . Transformacja opisana we wzorze (79) nie zmienia wartości współczynnika korelacji  $\tau$  Kendall'a i jest odbiciem

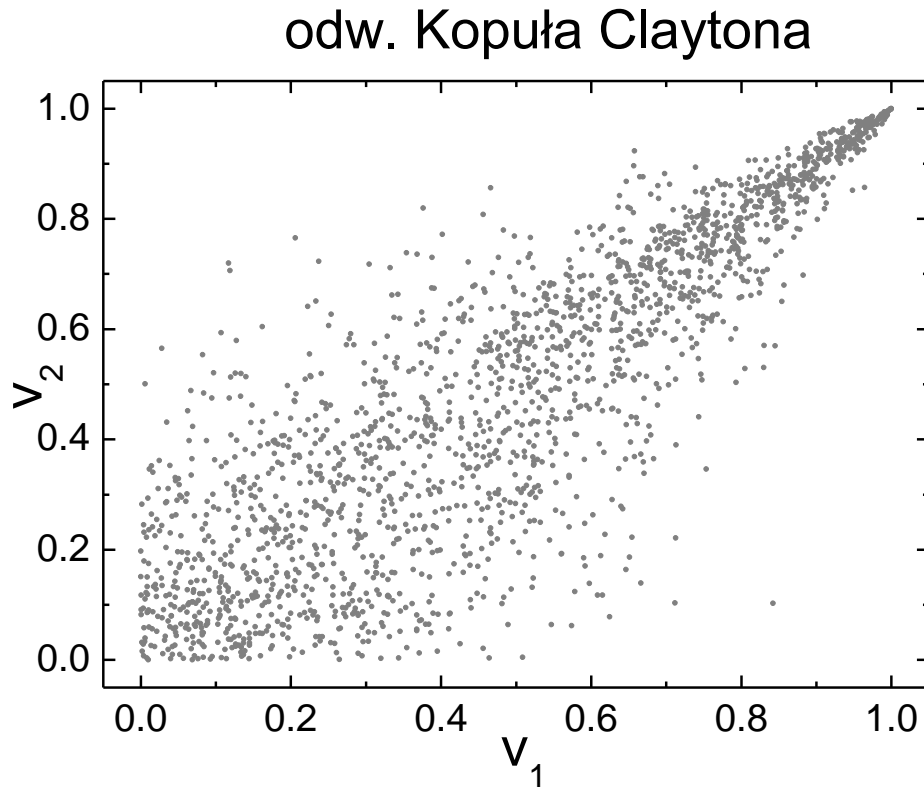


względem środka odcinka  $[0,1]$  dla obu zmiennych  $v_1$  oraz  $v_2$ . Z tego względu indeks dolej zależności ogonowej odwróconej Kopuły Gumbela jest równy indeksowi górnej zależności ogonowej Kopuły Gumbela, natomiast indeks górnej zależności ogonowej odwróconej Kopuły Claytona jest równy indeksowi dolnej zależności ogonowej Kopuły Claytona. Podobnie odwrócona Kopuła Gumbela nie posiada indeksu górnej zależności ogonowej, a odwrócona Kopuła Claytona nie posiada indeksu dolnej zależności ogonowej. W tabeli 5 przedstawiono indeksy zależności ogonowej dla odwróconej Kopuły Gumbela oraz Claytona. Ponieważ, jak pokazano, transformacja (79) nie wpływa na wielkość współczynnika korelacji  $\tau$  Kendall'a użytego do estymacji parametru  $\theta$  Kopuły, parametr ten dla Kopuły odwróconej będzie taki jak dla Kopuły pierwotnej.

Tabela 5. Indeksy zależności ogonowych dla odwróconej Kopuły Gumbela oraz Claytona.

Kopuła	$\lambda_l$	$\lambda_u$
odwrócona Gumbela $C_{\theta}^{RG}(v_1, v_2)$	$2 - 2^{\frac{1}{\theta}}$	0
odwrócona Claytona $C_{\theta}^{RC}(v_1, v_2)$	0	$2^{-\frac{1}{\theta}}$

Na rysunku 12 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez odwróconą Kopułę Claytona z parametrem  $\theta = 4.47$ , rysunek ten można porównać z danymi wygenerowanymi przez Kopułę Claytona - rysunek 10.



Rysunek 12. Dane wygenerowane przez odwróconą Kopułę Claytona.

Warto w tym miejscu zauważyć, że zależności ogonowe mają duże znaczenie w praktycznym wykorzystaniu Kopuł, ponieważ umożliwiają badanie jednoczesnego wystąpienia wartości ekstremalnych dla dwóch zmiennych, co okazuje się istotne w ocenie ryzyka jednoczesnego wystąpienia takich zdarzeń. Ze wzorów (76) oraz (78) wynika ponadto, że indeksy zależności ogonowej mają sens jedynie dla danych dodatnio skorelowanych.

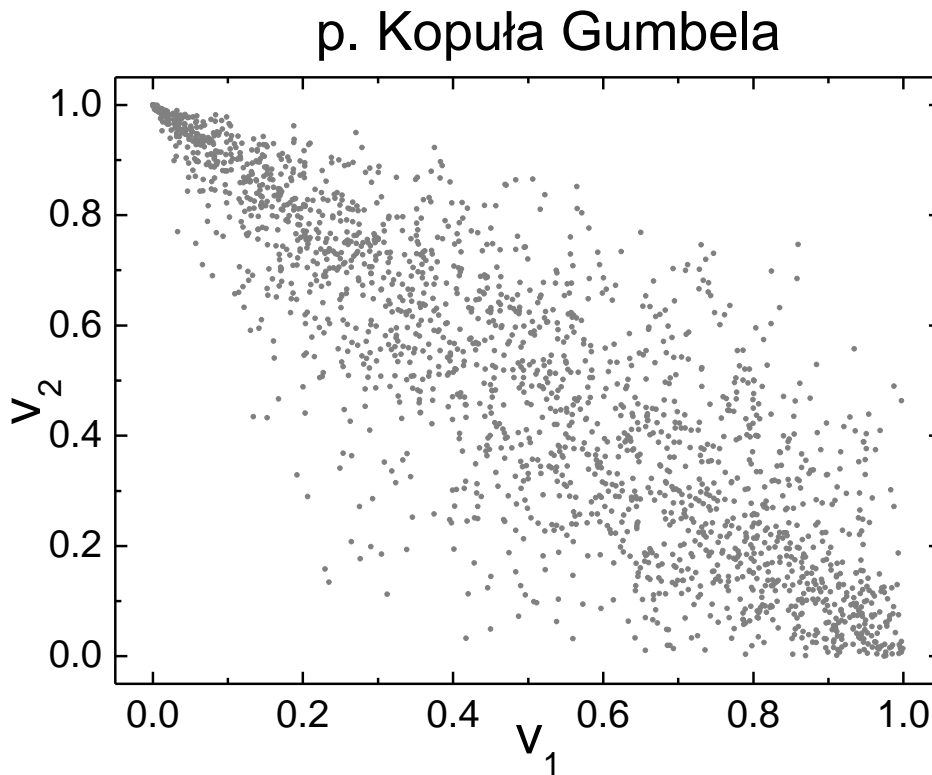
Porównując wzór (72) oraz warunek  $\theta \geq 1$  dla dwuwymiarowej Kopuły Gumbela, można wywnioskować, że Kopuła ta opisuje dane z  $\tau \geq 0$ , natomiast nie opisuje danych ujemnie skorelowanych z  $\tau < 0$ . Dlatego zaproponowano transformację (80) zmiennych pozwalającą na zdefiniowanie **przekręconej Kopuły Gumbela** -  $C_{\theta}^{rotG}(v_1, v_2)$  w celu zastosowania jej do analizy danych ujemnie skorelowanych:

$$\begin{cases} v_1 \rightarrow 1 - v_1, \\ v_2 \rightarrow v_2. \end{cases} \quad (80)$$

Transformacja ta jest transformacją odbicia względem środka jednorodnego odcinka  $[0,1]$ , na którym określona jest zmienna  $v_1$ . Ponadto zmienia ona współczynnik korelacji  $\tau$  Kendall'a w sposób następujący:

$$\tau \rightarrow -\tau. \quad (81)$$

Dlatego, aby wyznaczyć parametr  $\theta$  dla przekręconej Kopuły Gumbela, po wyliczeniu współczynnika  $\tau$  Kendall'a, należy zastosować wzór (81) a następnie relację pomiędzy  $\theta$  i  $\tau$  dla Kopuły Gumbela ujętą we wzorze (72). Na rysunku 13 przedstawiono dane w przestrzeni  $[0,1]^2$  wygenerowane przez Kopułę  $C_{\theta}^{rotG}(v_1, v_2)$  z parametrem  $\theta = 3.33$ . Rysunek 13 można porównać z rysunkiem 9 na którym przedstawiono dane wygenerowane przez Kopułę Gumbela.



Rysunek 13. Dane wygenerowane przez Kopułę  $C_{\theta}^{rotG}(v_1, v_2)$ .

Należy jeszcze zwrócić uwagę, że transformacja opisana we wzorze (80) może być przeprowadzona również w inny sposób, mianowicie według następującego wzoru:

$$\begin{cases} v_1 \rightarrow 1 - v_1 \\ v_2 \rightarrow v_2 \end{cases} \quad (82)$$

Który prowadzi również do relacji (81), jednak już do innej formy  $C_{\theta}^{rotG}(v_1, v_2)$  oraz innych danych wygenerowanych przez tą Kopułę - w tym przypadku na rysunku 13 należałoby dokonać odbicia względem środka odcinka  $[0,1]$  dla obu zmiennych  $v_1$  oraz  $v_2$ .

W następnych podrozdziałach zostaną przedstawione szczegółowo procedury estymacji parametrów poszczególnych Kopuł oraz generowania danych za pomocą tychże Kopuł.

### 3.4.4 Sposoby estymacji parametrów kształtu

W bieżącym podrozdziale zostaną wymienione metody estymacji parametrów funkcji Kopuły. Najbardziej uniwersalną metodą jest metoda największej wiarygodności. Wielowymiarową gęstość prawdopodobieństwa zmiennej losowej  $[x_1, \dots, x_i, \dots, x_n]$  można wyznaczyć z użyciem równania (65). Jeżeli jako  $[x_{1k}, \dots, x_{ik}, \dots, x_{nk}]_{k=1}^N$  zdefiniujemy zbiór danych empirycznych, otrzymujemy funkcję wiarygodności w postaci [1]:

$$l(\theta) = \sum_{k=1}^N \ln c(F_1(x_{1k}), \dots, F_i(x_{ik}), \dots, F_n(x_{nk})) + \sum_{k=1}^N \sum_{j=1}^n \ln f_j(x_{jk}), \quad (83)$$

gdzie  $\theta$  to zbiór parametrów zarówno funkcji Kopuły jak również funkcji rozkładów brzegowych. Następnie należy zmaksymalizować funkcję wiarygodności dla wszystkich parametrów zarówno Kopuły ( $C$ ) jak również rozkładów brzegowych -  $f_i(x_i)$ . Jednakże w pracy [49] zaproponowano procedurę polegającą na podziale funkcji wiarygodności na część odpowiadającą funkcji Kopuły -  $\sum_{k=1}^N \ln c(F_1(x_{1k}), \dots, F_i(x_{ik}), \dots, F_n(x_{nk}))$  oraz część odpowiadającą rozkładom brzegowym -  $\sum_{k=1}^N \sum_{j=1}^n \ln f_j(x_{jk})$ . Dzięki temu estymację parametrów Kopuły można przeprowadzić w dwóch krokach [1]:

- poprzez dopasowanie parametrów rozkładów brzegowych,
- poprzez dopasowanie parametrów Kopuły przy ustalonych parametrach rozkładów brzegowych.

W celu wyznaczenia parametrów niektórych funkcji Kopuły można stosować metody nieparametryczne. Wzór (72) pokazuje zależności parametru  $\theta$  podstawowych Kopuł

Archimedesowskich od współczynnika korelacji  $\tau$  Kendall'a, dlatego można go wykorzystać do estymacji parametru  $\theta$  [1]. Wartość współczynnika  $\tau$  Kendall'a nie zależy od jednowymiarowych rozkładów brzegowych, dlatego parametr Kopuły można dopasować niezależnie od parametrów rozkładów brzegowych.

W pracy [50] opisano metodę dopasowania parametrów Kopuły Studenta, polegającą na osobnym wyznaczeniu macierzy korelacji - wzór (42) - oraz osobnym wyznaczeniu parametru dyskretnego  $\nu$  opisującego ilość stopni swobody. W [50] pokazano również, że kolejne elementy macierzy korelacji  $\rho_{ij}$  można powiązać z współczynnikiem korelacji  $\tau$  Kendall'a według następującego wzoru:

$$\rho_{ij} = \sin\left(\frac{\pi}{2}\tau(X_i, X_j)\right). \quad (84)$$

Po wyborze oraz dopasowaniu rozkładów brzegowych, jak również wyznaczeniu współczynników macierzy korelacji, można wyznaczyć parametr  $\nu$  stosując metodę największej wiarygodności [50, 51]. W przypadku dwuwymiarowej Kopuły Studenta należy wyznaczyć dwa parametry: współczynnik korelacji  $\rho$ , pozwalający na wyznaczenie macierzy korelacji (44), oraz ilość stopni swobody  $\nu$ . Procedura dopasowania parametrów dwuwymiarowej Kopuły Studenta została przedstawiona poniżej w postaci następujących kroków:

- wybór i dopasowania rozkładów brzegowych  $f_1(x_1)$ , oraz  $f_2(x_2)$ ,
- transformacja danych do przestrzeni  $[0,1]^2$  za pomocą dystrybuanty  $v_i = F_i(x_i)$ ,
- wyznaczenie współczynnika korelacji  $\rho$  (można zastosować wzór 84),
- konstrukcja Kopuły Studenta oraz wyznaczenie parametru  $\nu$  metodą największej wiarygodności.

Opisane w tym podrozdziale procedury dopasowania parametrów funkcji Kopuły będą wykorzystane w następnych rozdziałach do analizy danych.

### 3.4.5 Sposoby generowania danych

W bieżącym podrozdziale omówię procedurę generowania danych do przestrzeni  $X_1, \dots, X_i, \dots, X_n$  z wykorzystaniem funkcji Kopuły. Procedura ta ma istotne zastosowanie, ponieważ pozwala symulować dane przy pomocy modelu zbudowanego przez funkcje Kopuły oraz dystrybuanty brzegowe - będzie ona wielokrotnie przeprowadzana na następnych rozdziałach. Generowanie danych polega na:

- generowaniu danych w przestrzeni  $[0,1]^n$  za pomocą funkcji Kopuły,
- transformacji danych do przestrzeni  $X_1, \dots, X_i, \dots, X_n$  z użyciem odwrotnych dystrybuant  $x_i = F_i^{-1}(v_i)$ .

Parametry funkcji Kopuły oraz dystrybuant są wyznaczane podczas dopasowania tych funkcji.

W tym miejscu zostanie omówiona procedura generowania danych z użyciem Kopuły Archimedesowskiej [1]. Załóżmy, że mamy Kopułę Archimedesowską o generatorze  $\phi(x)$ . Załóżmy ponadto, że mamy dodatnią zmienną losową  $\gamma$  oraz jej funkcję rozkładu prawdopodobieństwa  $G_\gamma$ . Transformata Laplace’a funkcji  $G$  jest opisana następująco:

$$\hat{G}(s) = \int_0^{+\infty} e^{-st} dG(t) \quad x \geq s. \quad (85)$$

Można udowodnić, że odwrotność transformaty Laplace’a daje w wyniku funkcję, która spełnia warunki na pełny generator Kopuły Archimedesowskiej [1]. Innymi słowy, istnieje taka funkcja  $\hat{G}(s)$ , której odwrotność  $\hat{G}^{-1}(x)$  odpowiada generatorowi badanej przez nas Kopuły Archimedesowskiej, czyli:

$$\hat{G}^{-1}(x) = \phi(x), \quad (86)$$

co można również zapisać jako:

$$\hat{G}(s) = \phi^{-1}(s). \quad (87)$$

Aby wygenerować dane przy pomocy Kopuły Archimedesowskiej należy przeprowadzić następującą procedurę [1]:

- odnaleźć rozkład prawdopodobieństwa  $G(t)$ , dla którego transformata Laplace’a jest równa odwrotności generatora Kopuły, np. w przypadku Kopuły Claytona jest to rozkład Gamma o parametrach  $\Gamma\left(1, \frac{1}{\theta}\right)$ , gdzie  $\theta$  to parametr Kopuły Claytona,
- wygenerować zmienną losową  $W$  z rozkładu  $G$ ,
- niezależnie wygenerować  $n$  niezależnych zmiennych losowych  $U_i$  z rozkładu jednorodnego  $[0,1]$ ,
- wyznaczyć zmienną losową  $v'_i = -\frac{\ln u_i}{w}$ ,

- przetransformować tę zmienną do przestrzeni  $[0,1]$  poprzez  $v_i = \hat{G}(v_i') = \phi^{-1}(v_i')$ ,
- otrzymane w ten sposób wartości zmiennej wielowymiarowej  $v_1, \dots, v_i, \dots, v_n$  są wielkościami wygenerowanymi przez odpowiednią Kopułę Archimedesowską.

Aby wygenerować dane za pomocą Kopuły Studenta można skorzystać z następującej procedury wynikającej z definicji Kopuły Studenta - wzór (50):

- należy wygenerować dane  $x_1, \dots, x_i, \dots, x_n$  z wielowymiarowego rozkładu Studenta  $t_{Rv}(x_1, \dots, x_i, \dots, x_n)$ ,
- dane  $x_1, \dots, x_i, \dots, x_n$  należy przetransformować za pomocą znormalizowanej jednowymiarowej dystrybuanty Studenta -  $t_v$  do przestrzeni  $[0,1]^n$  -  $v_i = t_v(x_i)$ .

Właściwości dwuwymiarowych funkcji Kopuły, przedstawione w bieżącym rozdziale, zostaną wykorzystane w kolejnych rozdziałach do analizy danych.

## 4. Modelowanie sygnałów geofizycznych i ekonofizycznych

### 4.1. Rozkłady prawdopodobieństwa i teoria wartości ekstremalnych

Ponieważ w pracy jest omawiana między innymi analiza danych ekstremalnych, dlatego warto przedstawić w tym miejscu parę faktów o statystyce tego typu danych.

Założmy, że mamy szereg czasowy będący zmienną losową  $z(t)$ . Jako szereg wartości maksymalnych  $x(t)$  rozumiemy wartości większe od pewnego progu  $z_0$  - znajdującego się w górnej części rozkładu prawdopodobieństwa zmiennej  $z(t)$ :

$$x(t) = z(t) \quad \text{jeżeli } z(t) \geq z_0. \quad (88)$$

Podobnie można badać wartości minimalne mniejsze od pewnego progu  $z'_0$  - znajdującego się w dolnej części rozkładu prawdopodobieństwa zmiennej  $z(t)$ :

$$x(t) = z(t) \quad \text{jeżeli } z(t) \leq z'_0. \quad (89)$$

Naturalnym wyborem podczas modelowania jednowymiarowych rozkładów wartości ekstremalnej jest uogólniony Rozkład Wartości Ekstremalnej (RWE) [52 - 58], który dla zmiennej stochastycznej  $x$  jest zdefiniowany następująco:

$$f(x, \xi) = g(x)^{1+\xi} e^{-t(x)}, \quad (90)$$

gdzie:

$$g(x, \xi) = \begin{cases} (1 + \xi x)^{-\frac{1}{\xi}} & \text{dla } \xi \neq 0 \\ e^{-x} & \text{dla } \xi = 0. \end{cases} \quad (91)$$



Parametr  $\xi$  pozwala rozróżnić trzy rodzaje rozkładów wartości ekstremalnej, których uogólnieniem jest RWE. Jeżeli  $\xi = 0$  mamy rozkład Gumbela [52, 53] opisany równaniem:

$$f(x) = e^{-e^x}. \quad (92)$$

Warto wspomnieć, że zgodnie z teorią wartości ekstremalnych, jeżeli zmienna losowa  $z$  ma rozkład Gaussa lub eksponentialny, to wartości ekstremalne mają rozkład Gumbela [52, 53].

Dalsza dyskusja zostanie przeprowadzona dla przykładu opisanego wzorem (88). W przypadku opisanym wzorem (89) rozkład Fréchéta zajmuje miejsce rozkładu Weibulla i vice versa. Rozważając teraz przypadek  $\xi > 0$  oraz definiując  $\xi = c^{-1}$  otrzymujemy rozkład Fréchéta [54] określony poprzez:

$$f(x, c) = \begin{cases} c(x)^{c-1} e^{-(x)^c} & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0. \end{cases} \quad (93)$$

Jeżeli dane pierwotne ( $z$ ) mają rozkład potęgowy, to ich wartości maksymalne mają rozkład Fréchéta [54]. Ostatecznie rozważając przypadek  $\xi < 0$  oraz definiując  $\xi = -c^{-1}$  otrzymujemy rozkład Weibulla [55] określony poprzez:

$$f(x, c) = \begin{cases} c(-x)^{c-1} e^{-(-x)^c} & \text{dla } x < 0 \\ 0 & \text{dla } x \geq 0. \end{cases} \quad (94)$$

Do rozkładu Weibulla dążą wartości ekstremalne zmiennej losowej mające ograniczenie górne - np. zmiennej pochodzącej z rozkładu jednorodnego.

Aby wprowadzić parametry skali do rozkładu (RWE), oraz pozostałych rozkładów, które zostaną omówione w następnym paragrafie, zastosowano standardową zmianę zmiennych:

$$x \rightarrow \frac{x - \mu}{a}, \quad (95)$$

oraz:

$$f(x) \rightarrow \frac{1}{a} f\left(\frac{x - \mu}{a}\right). \quad (96)$$

Reasumując, Uogólniony Rozkład Wartości Ekstremalnej, który zawiera w sobie rozkłady Gumbela, Weibulla i Fréchéta, umożliwia analizę różnego rodzaju danych ekstremalnych pochodzących z różnych rozkładów pierwotnych. Należy jednak zwrócić w tym miejscu uwagę na fakt, że auto-korelacja danych ma wpływ na rozkłady prawdopodobieństwa wartości ekstremalnych [59], co zostanie omówione w następnym podrozdziale.

#### 4.2. Inne rozkłady, dane ekstremalne a auto-korelacje

Jak już wspomniano auto-korelacje wpływają na rozkłady prawdopodobieństwa wartości ekstremalnych, powodując odejście od teorii wartości ekstremalnych oraz od rozkładu (RWE) [59]. Dlatego zostanie omówionych kilka innych ogólnie stosowanych rozkładów prawdopodobieństwa, aby podczas analizy danych wybrać optymalny rozkład z użyciem odpowiedniego testu statystycznego. Omówione rozkłady nie są związane teorią wartości ekstremalnej i dlatego można je spróbować zastosować zarówno do analizy danych pierwotnych typu  $z(t)$ , jak również danych ekstremalnych typu  $x(t)$

Pierwszym rozkładem jest dość uniwersalny rozkład Johnsona  $S_U$  [60] ( $RJS_U$ ). Znajduje on zastosowanie w analizie ekstremalnych danych finansowych oraz modelach typu „VaR”, co zostanie pokazane w następnym rozdziale. Okazuje się jednak, że rozkład ten jest równie użyteczny do badania danych geofizycznych. Gęstość prawdopodobieństwa rozkładu Johnsona  $S_U$  jest opisana poprzez wzór:

$$f(x, \zeta, \nu) = \frac{\nu}{\sqrt{x^2+1}} \Phi\left(\zeta + \nu \ln(x^2 + \sqrt{x^2 + 1})\right), \quad (97)$$

gdzie  $\Phi$  to gęstość prawdopodobieństwa znormalizowanego rozkładu Gaussa, natomiast parametry  $\nu$  oraz  $\zeta$  są związane z wyższymi momentami rozkładu. Rozkład Johnsona  $S_U$  jest więc pewnym rozszerzeniem rozkładu Gaussa umożliwiającym badanie grup danych posiadających różną asymetrię i kurtozę.

Warto w tym miejscu wspomnieć również o rozkładzie Logistycznym (RL), używanym do analizy danych hydrologicznych [61, 62], którymi jest część danych badanych w ramach realizacji pracy doktorskiej, to znaczy:

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}. \quad (98)$$

Do analizy danych ekstremalnych posiadających różne auto-korelacje zastosowano następującą procedurę. W pierwszej kolejności dopasowano rozkład (RWE) oraz zastosowano odpowiedni test statystyczny do oceny dopasowania. Następnie, spróbowano dopasować inne rozkłady ( $RJS_U$ , RL) celem wyboru najlepszego modelu. Jak się okazało, dla niektórych danych posiadających auto-korelacje, rozkłady inne niż (RWE) reprezentowały model lepszy od tego, gdy używano (RWE). W przypadku danych pierwotnych - typu  $z(t)$  - naturalnym wyborem jest rozkład Gaussa (RG) zgodnie z Centralnym Twierdzeniem Granicznym. Jednak wiadomo, że wystąpienie auto-korelacji powoduje odejście od Centralnego Twierdzenia Granicznego [1]. Dlatego w przypadku wykrycia auto-korelacji warto również zbadać ( $RJS_U$ ), który jest rozszerzeniem (RG).

### **4.3. Auto-korelacje a wielowymiarowa statystyka - zastosowanie wykładnika Hursta w procedurze wyboru funkcji Kopuły**

W poprzednich rozdziałach zauważono, że auto-korelacje mają wpływ na jednowymiarowe rozkłady prawdopodobieństwa, sugerując dla danych ekstremalnych różne od (RWE) rozkłady prawdopodobieństwa. Warto w tym miejscu zwrócić uwagę, że Kopuła Gumbela należy do grupy Kopuł wartości ekstremalnych i jest używana do analizy takich danych [63]. Dlatego analogicznie jak w przypadku jednowymiarowych rozkładów prawdopodobieństwa należało użyć również innych funkcji Kopuł do badania danych ekstremalnych posiadających auto-korelacje - np. w przypadku danych maksymalnych oprócz Kopuły Gumbela stosowano odwróconą Kopułę Claytona.

Warto zauważyć, że ogólna auto-korelacja danych prowadzi do tworzenia się klastrów danych ekstremalnych [64] - np. po wystąpieniu bardzo dużej wartości wzrasta prawdopodobieństwo, że następna wartość będzie również bardzo duża. Zwiększa to lokalnie ryzyko wystąpienia takiej wartości, sugerując lokalne zastosowanie funkcji Kopuły o wyższym indeksie zależności ogonowej. Dodatkowo nadmienimy, że w przypadku danych finansowych, gdzie występuje zmienna auto-korelacja, pojawienie się ujemnej auto-korelacji sprzyja występowaniu zjawisk ekstremalnych. Dlatego sygnał anty-korelacji również lokalnie sugeruje zastosowanie funkcji Kopuły o wyższym indeksie zależności ogonowej. Z tego powodu można wykorzystać wiedzę o występowaniu auto-korelacji w procedurze wyboru odpowiedniej funkcji Kopuły, co zostanie użyte wielokrotnie w kolejnych rozdziałach.

## 5. Analiza danych empirycznych

### 5.1. Badanie izotopów w rdzeniach arktycznych

Model obiegu pierwiastków w przyrodzie jest złożony i ma wiele zmiennych. Przyglądając się historii koncentracji pierwiastków można wywnioskować, że model ich obiegu zawiera elementy stochastyczne oraz deterministyczne, co odzwierciedla różny poziom korelacji pomiędzy różnymi koncentracjami oraz różny poziom auto-korelacji poszczególnych koncentracji. Taka dynamika może okazać się analogiczna do dynamiki danych finansowych.

Badaniom zostały poddane szeregi czasowe odzwierciedlające historię koncentracji radioizotopów oraz jonów w rejonie bieguna południowego. W tym celu zostały zbadane dane zarejestrowane w rdzeniach lodowych otrzymanych podczas odwiertów wykonanych w stacji badawczej w „Taylor Dome” na Antarktydzie [65 - 69], na wzniesieniu lodowym określonym przez współrzędne  $77^{\circ}47,7'S$   $158^{\circ}43,1'E$ , na wysokości około 2400 m n.p.m. „Taylor Dome” znajduje się około 150 km od zachodniego brzegu Morza Rossa i jest przestrzenią, w której następuje akumulacja lodu [65, 67] - lokalizację tą przedstawiono na rysunku 14. Rdzenie uzyskane w „Taylor Dome” dostarczają szczegółowych zapisów historycznych warunków klimatycznych panujących na terenach znajdujących się na zachód od Morza Rossa [67], w tym zawierają informację o koncentracji izotopów kosmogenicznych oraz kilku innych jonów.



Rysunek 14. Lokalizacja „Taylor Dome” – źródło [70].

Badaniu poddano koncentracje następujących izotopów oraz jonów:

- izotop  $^{10}\text{Be}$  [66, 67],
- izotop  $^{18}\text{O}$  [67, 68],
- izotop  $^2\text{D}$  [66, 67, 69]
- jony  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$  [65, 67, 71].

Koncentracja jonów oraz izotopów dostępna jest w różnych jednostkach. Koncentrację izotopu  $^{10}\text{Be}$  podano w ilości atomów na miligram, koncentrację jonów  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$  w ppb - ilości atomów na miliard ( $10^9$ ) - natomiast koncentrację izotopów  $^{18}\text{O}$  oraz  $^2\text{D}$  w stosunku do zestandaryzowanej czystej wody oceanicznej (SMOW) według poniższych wzorów:

$$\delta^{18}\text{O}_{\text{próbki}} = \left( \frac{\left( \frac{^{18}\text{O}}{^{16}\text{O}} \right)_{\text{próbki}}}{\left( \frac{^{18}\text{O}}{^{16}\text{O}} \right)_{\text{SMOW}}} - 1 \right) \cdot 1000 \text{ ‰} , \quad (99)$$

$$\delta^2\text{D}_{\text{próbki}} = \left( \frac{\left( \frac{^2\text{D}}{^1\text{H}} \right)_{\text{próbki}}}{\left( \frac{^2\text{D}}{^1\text{H}} \right)_{\text{SMOW}}} - 1 \right) \cdot 1000 \text{ ‰} , \quad (100)$$

gdzie  $\left(\frac{^{18}O}{^{16}O}\right)_{SMOW} = 2005.20 \text{ ppm}$  (cząstek na milion -  $10^6$ ), natomiast  $\left(\frac{^2D}{^1H}\right)_{SMOW} = 155.76 \text{ ppm}$ .

Koncentracja  $^{18}O$  oraz  $^2D$  w rdzeniach lodowych jest mniejsza od poziomów referencyjnych, dlatego wartości  $\delta^{18}O_{próbki}$  oraz  $\delta^2D_{próbki}$  są ujemne.

Datowanie próbek zostało przeprowadzone z użyciem skali czasowej st9810 [67], która określa relację pomiędzy głębokością, na jakiej znajduje się próbka oraz wiekiem próbki. Skala ta została wyznaczona z użyciem kombinacji analizy geochemicznej oraz geofizycznego modelu przepływu lodu [67]:

- dla okresu do 11.7 tys. lat - datowanie przeprowadzono na podstawie modelu przepływu oraz narastania lodu,
- dla okresu 11.7 – 20 tys. lat - datowanie przeprowadzono na podstawie koncentracji  $CH_4$  oraz  $\delta^{18}O_{atm}$ ,
- dla okresu powyżej 20 tys. lat - datowanie przeprowadzono z użyciem zaawansowanego modelu wykorzystującego koncentrację atmosferycznych gazów, dodatkowo wykorzystano dwie anomalie koncentracji  $^{10}Be$ , które wystąpiły 37 tys. lat oraz 64 tys. lat temu.

Niepewność skali czasowej jest następująca [67]:

- dla okresu do 20 tys. lat - mniej niż 5%,
- wraz ze wzrostem głębokości niepewność rośnie od 5% dla 36 tys. lat do około 10% dla 130 tys. lat,
- natomiast dla okresu powyżej 130 tys. lat niepewność jest jeszcze większa [67].

Po omówieniu rodzaju i pochodzeniu danych, przejdziemy w kolejnym podrozdziale do przedstawienia fizycznego modelu obiegu pierwiastków w atmosferze.

#### 5.1.1 Obieg pierwiastków w atmosferze

Aby dobrze opisać obieg jonów oraz izotopów w atmosferze należy je podzielić na kilka grup. Pierwsza grupa to izotopy kosmogeniczne, których obecność w atmosferze wynika z kolizji promieniowania słonecznego z cząsteczkami znajdującymi się w górnej części atmosfery ziemskiej. Ilość wyprodukowanych izotopów kosmogenicznych zależy od aktywności Słońca, natężenia promieniowania kosmicznego oraz natężenia ziemskiego pola magnetycznego [72, 73]. Jednym z takich izotopów jest beryl  $^{10}Be$ . Izotop ten pozostaje w atmosferze ziemskiej przez około 1-2 lata a następnie trafia na powierzchnię Ziemi, dzięki czemu można go znaleźć w badanych rdzeniach

lodowych. Koncentracja izotopu  $^{10}\text{Be}$  jest zbliżona w różnych miejscach na Ziemi, dlatego znajduje ona zastosowanie w wyznaczaniu skali czasowych określających czas pochodzenia próbki w funkcji głębokości odwiertu, kiedy bada się tzw. anomalie - okresy, w których produkcja izotopu  $^{10}\text{Be}$  była znacznie zwiększona. Pomimo, że koncentrację izotopu  $^{10}\text{Be}$  w atmosferze kształtuje aktywność Słońca [73], będąca czynnikiem deterministycznym, to należy jednak pamiętać, że okres cyklu słonecznego wynoszący 11 lat jest nieporównywalnie krótki w stosunku do szerokości skali czasowych, w jakich badane są próbki, a wpływ długoterminowych zmian w aktywności Słońca na koncentrację izotopu  $^{10}\text{Be}$  jest niewielki [74]. Ogólnie wiadomo, że istotny wpływ na koncentrację w atmosferze izotopu  $^{10}\text{Be}$  ma natężenie pola magnetycznego Ziemi. Okazuje się jednak, że cyrkulacyjny obieg powietrza w okolicy bieguna powoduje, iż w obszarze tym w atmosferze znajduje się głównie izotop  $^{10}\text{Be}$ , który powstał lokalnie, co zmniejsza wpływ zmian natężenia pola magnetycznego Ziemi na koncentrację izotopu  $^{10}\text{Be}$  w tym miejscu [67, 75]. Oprócz wpływu czynników deterministycznych na koncentrację  $^{10}\text{Be}$  w rdzeniach lodowych mają wpływ również inne czynniki o charakterze stochastycznym, np. natężenie opadów lub akumulacja śniegu [67, 76].

Kolejnym obiektem badań jest para izotopów  $^2\text{D}$  oraz  $^{18}\text{O}$ , które występują naturalnie na Ziemi a ich udział w wodzie morskiej wynosi odpowiednio 0.015% oraz 0.20%. Koncentracja tlenu 18 lub deuteru może być wykorzystywana do analizy temperatury na Ziemi, ponieważ ich koncentracja w rejonach polarnych rośnie wraz ze wzrostem temperatury [72]. Wynika to z faktu, że maksymalna zawartość pary wodnej w powietrzu maleje wraz ze wzrostem temperatury. Jeżeli rozważymy parę wodną, która powstaje w cieplejszych rejonach oceanicznych i przemieszcza się w kierunku chłodniejszych biegunów, maksymalna wilgotność powietrza maleje wraz z temperaturą, co powoduje kondensację części pary wodnej. Cząsteczki wody zawierające cięższe izotopy ( $^2\text{D}$  lub  $^{18}\text{O}$ ) ulegają kondensacji z większym prawdopodobieństwem. Dlatego para wodna, która dociera w rejony polarne oraz śnieg padający w tych rejonach są uboższe w izotopy ( $^2\text{D}$  lub  $^{18}\text{O}$ ) niż woda padająca w cieplejszych rejonach Ziemi. Z drugiej strony im niższa temperatura, tym mniejsza maksymalna wilgotność powietrza, które dociera w rejony biegunów, a co za tym idzie mniejsza procentowa zawartość izotopów ( $^2\text{D}$  lub  $^{18}\text{O}$ ) w śniegu. Z tego powodu oba izotopy mają bardzo podobny mechanizm obiegu w cząsteczkach wody w atmosferze, silnie zależny od temperatury na Ziemi, która jest czynnikiem globalnym mającym duży wpływ na koncentrację obu izotopów oraz silną korelację. Dlatego zarówno koncentrację izotopu deuteru  $\delta^2D$  oraz tlenu  $\delta^{18}O$  wykorzystuje się w celu symulowania temperatury, która występowała w przeszłości, w danym miejscu, poprzez relacje liniowe ujęte w poniższych wzorach [67]:

$$\alpha_O = \frac{\Delta\delta^{18}O}{\Delta T}, \quad (101)$$

$$\alpha_D = \frac{\Delta\delta^2 D}{\Delta T}, \quad (102)$$

gdzie  $\Delta T$  to zmiana temperatury powierzchni Ziemi w punkcie pobierania próbek - w naszym przypadku „Taylor Dome”, natomiast  $\alpha_O$  oraz  $\alpha_D$  to współczynniki liniowe, spełniające deterministyczną zależność  $\alpha_O \approx \frac{1}{8}\alpha_D$ . Możemy więc założyć, że opisany czynnik globalny będzie miał przewagę nad czynnikami lokalnymi kształtującymi osobno koncentrację tlenu 18 i deuteru, co będzie się uwidaczniać w wysokiej, prawie deterministycznej, korelacji.

Omówmy jeszcze pokrótce obieg jonów, które są transportowane w powietrzu w formie aerozoli i mają różne pochodzenie; morskie lub lądowe, to znaczy:  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ . Wśród badanych jonów możemy wyróżnić jon pochodzenia morskiego -  $\text{Na}^+$ , oraz jon pochodzenia lądowego -  $\text{Ca}^{2+}$  [67]. W pracy [77] zasugerowano, iż pozostałe jony, np.  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , mogą być częściowo pochodzenia lądowego, gdzie  $\text{K}^+$  pochodzi z itów, natomiast  $\text{Ca}^{2+}$  oraz  $\text{Mg}^{2+}$  z osadów szelfów kontynentalnych. Mogą one również częściowo pochodzić z morza, a dokładnie z soli morskiej. Koncentracja jonów w typowej wodzie morskiej została podana w tabeli 6.

Tabela 6. Koncentracja jonów w typowej wodzie morskiej [78]

Jony	Koncentracja [g/l]
$\text{Cl}^-$	19.0
$\text{Na}^+$	10.6
$\text{Mg}^{2+}$	1.3
$\text{Ca}^{2+}$	0.4
$\text{K}^+$	0.4

Po przeliczeniu danych z tabeli 6, z użyciem odpowiednich mas atomowych, można zestawzić stosunek ilości cząstek poszczególnych jonów w stosunku do ilości cząstek  $\text{Na}^+$  znajdujących się w wodzie morskiej - tabela 7.



Tabela 7. Koncentracja poszczególnych jonów znajdujących się w wodzie morskiej, wyrażona w ilości cząsteczek, w stosunku do ilości cząsteczek  $\text{Na}^+$ .

Jony	Ilość cząstek / $\text{Na}^+$
$\text{Cl}^-$	1.16
$\text{Mg}^{2+}$	0.116
$\text{Ca}^{2+}$	0.021
$\text{K}^+$	0.022

W tym miejscu zostanie zaproponowana przybliżona procedura mająca na celu wyznaczenie pochodzenia koncentracji pozostałych jonów. Zakłada się, że jon  $\text{Na}^+$  jest pochodzenia morskiego a koncentracja jonów w wodzie morskiej przez ostatnie 10 tys. lat była w przybliżeniu stała, natomiast obieg różnych jonów pochodzenia morskiego miał podobny charakter, gdyż dostawały się one do atmosfery głównie lokalnie, z soli morskiej. Średnia koncentracja jonu  $\text{Na}^+$  przez ostatnie 10 tys. lat wynosiła około 10.1 ppb. Jeśli założymy, że obieg pozostałych jonów jest podobny do obiegu sodu, ich przewidywaną koncentrację, w części pochodzącej z morza, można wyliczyć na podstawie tabeli 7 - wyniki wyliczeń przedstawiono w tabeli 8.

Tabela 8. Sugerowana średnia koncentracja jonów na podstawie koncentracji jonu  $\text{Na}^+$  równiej 10.1 ppb.

Jony	Sugerowana średnia koncentracja [ppb]
$\text{Cl}^-$	11.7
$\text{Mg}^{2+}$	1.17
$\text{Ca}^{2+}$	0.21
$\text{K}^+$	0.22

W tabeli 9 pokazano faktyczną średnią koncentrację wymienionych jonów przez okres ostatnich 10 tys. lat.

Tabela 9. Faktyczna koncentracja jonów za okres ostatnich 10 tys. lat.

Jony	Faktyczna średnia koncentracja [ppb]
Cl <sup>-</sup>	38.9
Mg <sup>2+</sup>	1.53
Ca <sup>2+</sup>	0.99
K <sup>+</sup>	0.76

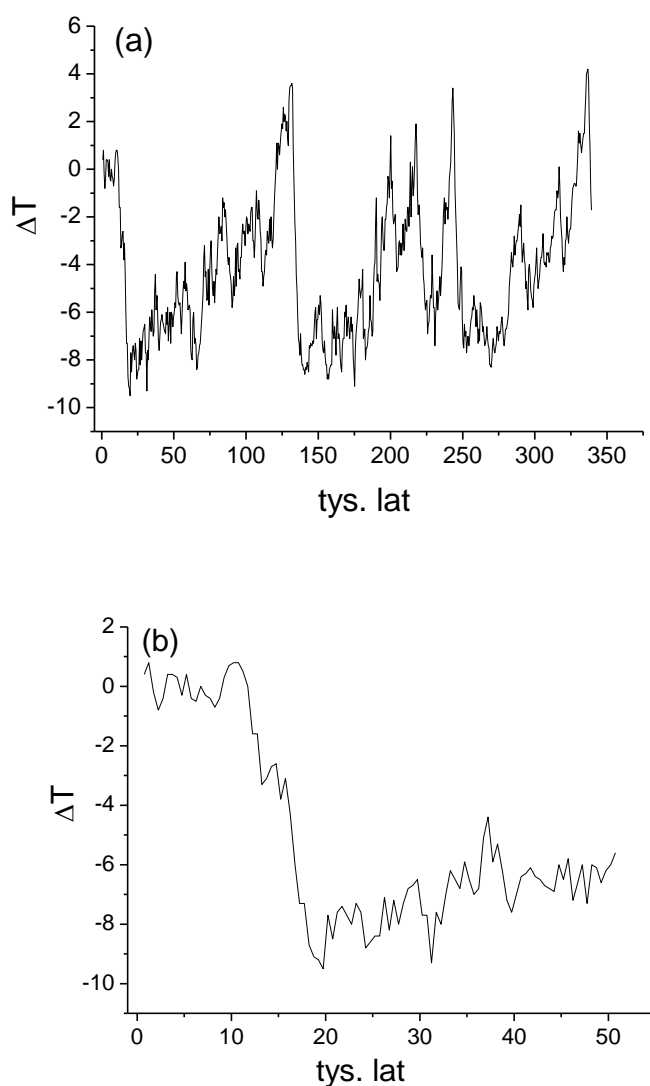
Porównując tabelę 8 i 9 można stwierdzić, że średnią koncentrację najbliższą sugerowanej na podstawie składu wody morskiej ma magnez, dlatego można założyć, iż jon ten pochodzi głównie z morza. Należy jednak pamiętać, iż w pewnej mniejszej części pochodzi on również z lądu [77]. W przypadku wapnia, który jest pochodzenia lądowego [67] różnica jest największa - faktyczna koncentracja jest prawie 5 razy większa niż ta sugerowana przez zawartość wapnia w wodzie morskiej. Dla pozostałych dwóch jonów - Cl<sup>-</sup> oraz Mg<sup>2+</sup> występuje sytuacja pośrednia, sugerująca ich częściowo morskie oraz częściowo lądowe pochodzenie, jednak ich faktyczna koncentracja jest ponad 3 razy większa, co sugeruje głównie pochodzenie lądowe. Dlatego założono następujące pochodzenie jonów:

- sód - morskie,
- magnez - głównie morskie ze składową lądową - z osadów szelfów kontynentalnych,
- potas i chlor - głównie lądowe (potas z ilów) ze składową morską,
- wapń - lądowe - z szelfów kontynentalnych.

W dalszej części pracy będziemy szukać argumentów na potwierdzenie takiego pochodzenia jonów oraz wpływu czynników globalnych i lokalnych na obieg jonów. Warto przywołać wspólne mechanizmy cyrkulacji, które odpowiadają za czynniki globalne. Globalnym czynnikiem kształtującym koncentrację jonów nad Antarktydą jest siła układów cyklonicznych odpowiedzialnych za transport aerozoli w kierunku południowym [67]. Stwierdzono ponadto, iż większe odkładanie się wszystkich jonów występuje ogólnie podczas zimniejszych okresów [67]. W mechanizmie cyrkulacji występują również czynniki lokalne. Ogólnie wzrost koncentracji jonów pochodzenia morskiego wynika ze wzrostu intensywności wiatrów wiejących nad obszarami morskimi, natomiast wzrost koncentracji jonów pochodzenia lądowego jest związany ze wzrostem intensywności wiatrów wiejących nad obszarami lądowymi [67]. W przypadku jonów pochodzenia lądowego kolejne lokalne czynniki są związane z różnym pochodzeniem lądowym. Jony wapnia oraz jony magnezu (w mniejszej części

która pochodzi z lądu) pochodzą z osadów szelfów kontynentalnych [67, 77], natomiast jony potasu pochodzą z ilów [77]. Dlatego warto porównać koncentrację wszystkich jonów z koncentracją jonów pochodzenia morskiego -  $\text{Na}^+$  oraz pochodzenia lądowego -  $\text{Ca}^{2+}$ .

Można więc przyjrzeć się historii klimatu na Ziemi poprzez rekonstrukcję temperatury z danych zarejestrowanych w „Taylor Dome”, jak również w innych lokalizacjach [67]. Możemy stwierdzić, że okres ostatnich 10-11 tysięcy lat był okresem ciepłym zwanym Holocenem - zakłada się, że okres Holocenu rozpoczął się około 11.7 tysięcy lat temu. Okres Holocenu był poprzedzany przez okres Pleistocenu, którego końcowa część była okresem zimnym, ostatnim okresem zlodowacenia. Okres ten rozpoczął się około 110 tys. lat temu i zakończył się około 12 tys. lat temu. Na rysunku 15 można zobaczyć rekonstrukcję temperatury dla stacji badawczej „Dome Fuji” na Antarktydzie [79].

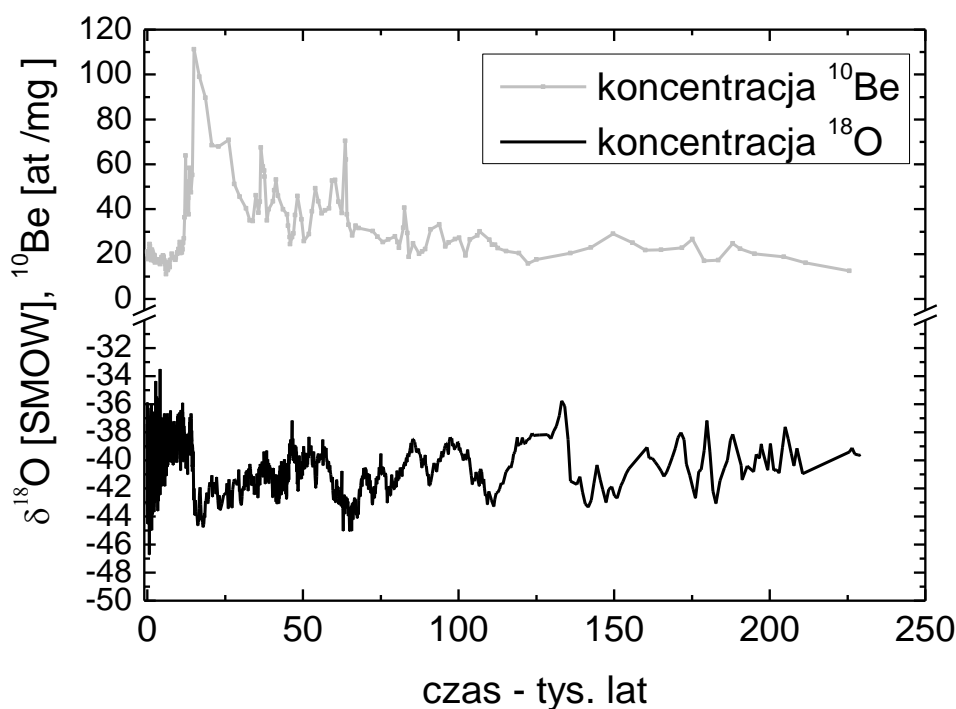


Rysunek 15. Rekonstrukcja średniej temperatury  $\Delta T$  - w stosunku do teraźniejszej przeprowadzona dla „Dome Fuji” na Antarktydzie [79]. Wykresy uzyskano w (a) - długiej, (b) - krótkiej skali czasowej.

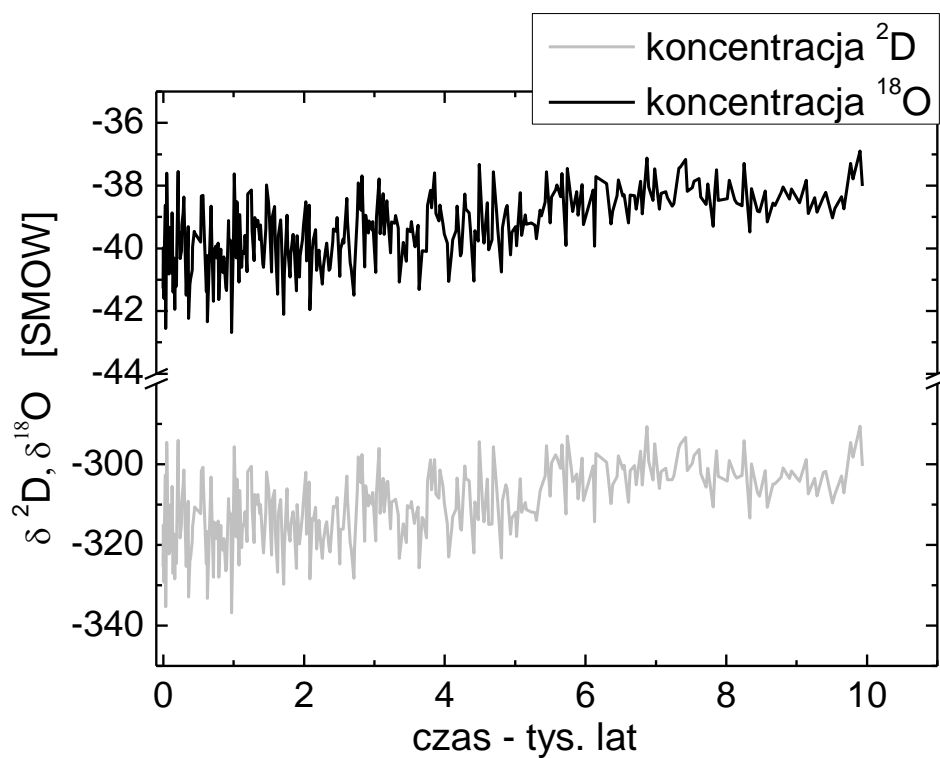
Okres ostatnich 10 tysięcy lat cechował się stosunkowo stabilną temperaturą, co powinno wpływać na stabilny układ wiatrów oraz stabilny obieg jonów i izotopów w atmosferze podlegający jednemu modelowi. Pozwala to na analizę czynników globalnych, wspólnych dla obiegu określonej grupy izotopów lub jonów. Dlatego do badań koncentracji pierwiastków w rdzeniach lodowych wybrano ten okres.

Wyjątkowo analizę korelacji pomiędzy danymi kosmogenicznymi - koncentracją  $^{10}\text{Be}$  oraz danymi silnie związanymi z temperaturą na Ziemi - koncentracją  $^{18}\text{O}$  - przeprowadzono dla okresu o długości 225 tys. lat - zawierającego zarówno okresy ciepłe jak i zimne. Dzięki temu podjęto próbę badania korelacji pomiędzy czynnikami związanymi z koncentracją  $^{10}\text{Be}$  a temperaturą silnie związaną z koncentracją  $^{18}\text{O}$  w okresach zarówno zimnych jak i ciepłych.

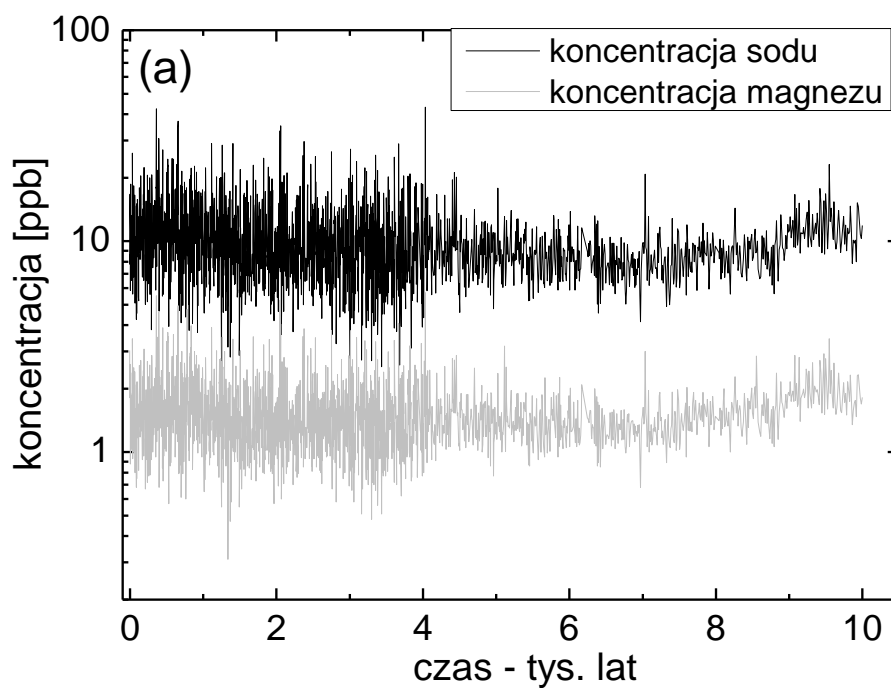
Na rysunku 16 pokazano koncentrację  $^{10}\text{Be}$  oraz  $^{18}\text{O}$ . Pobieźna analiza samego rysunku nie pozwala na stwierdzenie o występowaniu korelacji pomiędzy zmiennymi.



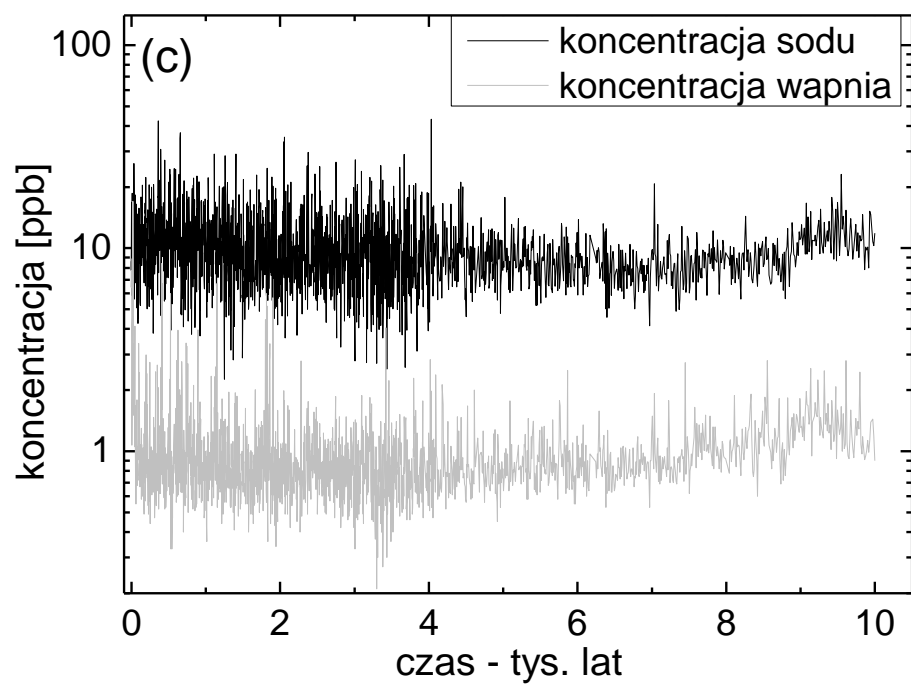
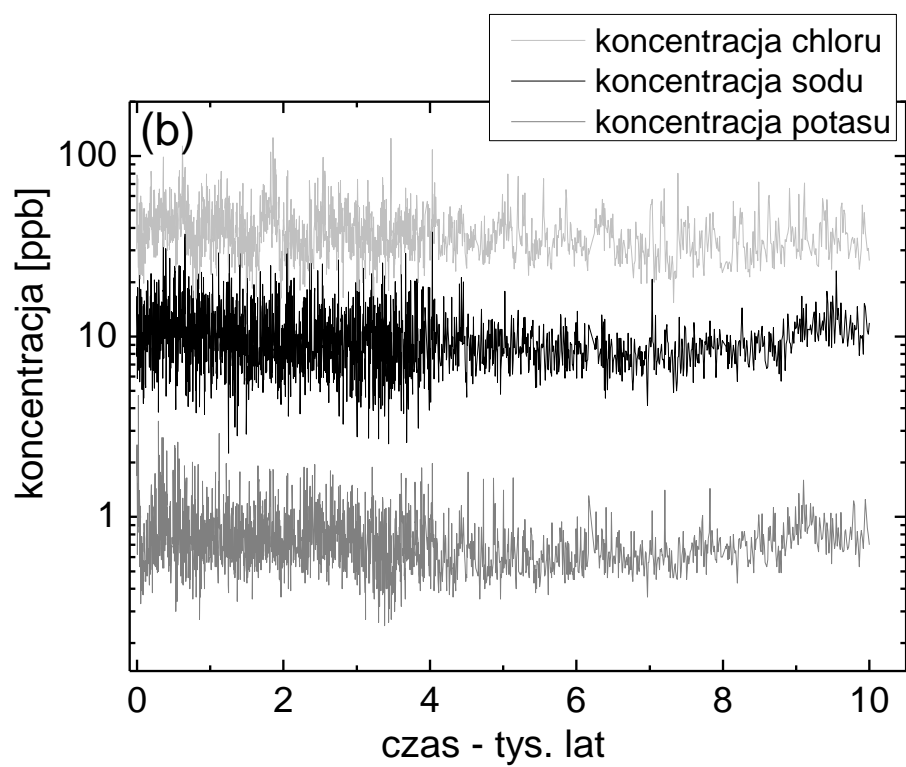
Rysunek 16. Koncentracja  $^{10}\text{Be}$  oraz  $^{18}\text{O}$  - ostatnie 225 tys. lat. Na rysunku ujęto czas historyczny w odniesieniu do teraźniejszości.



Rysunek 17. Koncentracja  $^2\text{D}$  oraz  $^{18}\text{O}$  - ostatnie 10 tys. lat, czas historyczny.



Rysunek 18. Koncentracja jonów: (a) -  $\text{Na}^+$ ,  $\text{Mg}^{2+}$  - ostatnie 10 tys. lat, czas historyczny.



Rysunek 18 kont. Koncentracja jonów: (b) - $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{K}^+$ ; (c) - $\text{Na}^+$ ,  $\text{Ca}^{2+}$  - ostatnie 10 tys. lat, czas historyczny.

Na rysunku 17 pokazano koncentrację  $^2\text{D}$  oraz  $^{18}\text{O}$  za ostatnie 10 tys. lat. Analizując rysunek można zaobserwować dobrą korelację pomiędzy zmiennymi. Na rysunku 18 przedstawiono natomiast koncentrację jonów  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  oraz  $\text{Cl}^-$  za ostatnie 10 tys. lat -  $K^{\text{Na}}(t)$ ,  $K^{\text{K}}(t)$ ,  $K^{\text{Ca}}(t)$ ,  $K^{\text{Mg}}(t)$ ,  $K^{\text{Cl}}(t)$ . Z rysunków można wyczytać stosunkowo dobrą korelację pomiędzy koncentracją  $\text{Na}^+$  oraz  $\text{Mg}^{2+}$ . W następnych rozdziałach zostanie przedstawiona analiza auto-korelacji danych oraz analiza korelacji pomiędzy danymi przedstawionymi na rysunkach 16, 17 oraz 18.

### 5.1.2 Badania auto-korelacji

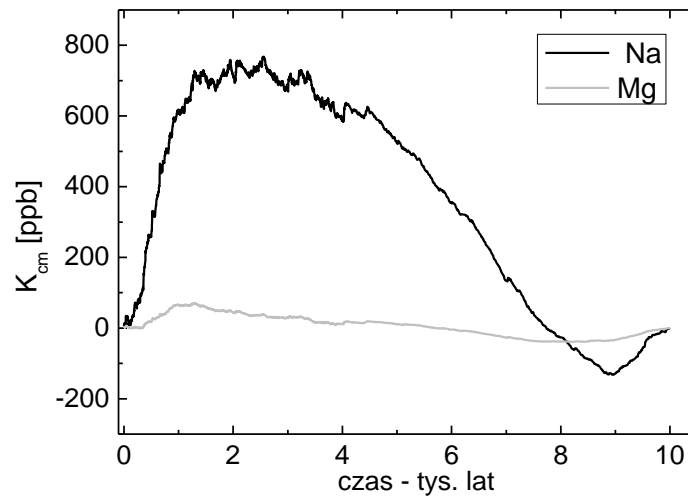
Badania przeprowadzono dla szeregów czasowych za ostatnie 10 tys. lat, gdzie ze względu na stabilną temperaturę można założyć stabilny układ wiatrów oraz stabilny obieg jonów i izotopów w atmosferze. Badanie auto-korelacji może dostarczyć ważnych informacji na temat szeregów czasowych. Różne pochodzenie jonów  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  oraz  $\text{Cl}^-$  może sugerować różne auto-korelacje oraz wpływ czynników deterministycznych i stochastycznych. Jeżeli koncentrację jonów w czasie  $t_n$  zapiszemy jako  $K(t_n)$ , zgodnie z modelem zaproponowanym w rozdziale 2, zakładając, że jest ona częściowo stochastyczna i częściowo deterministyczna, to pozwala to zaadoptować wzór (31) i zapisać go w postaci (103):

$$K(t_n) = K_0 + W + k(K(t_{n+1}), K(t_{n+2}), \dots, K(t_{n+m})), \quad (103)$$

gdzie użyto oznaczeń analogicznych jak we wzorze (31):  $W$  to czynnik losowy,  $k$  odpowiada za auto-korelacje,  $t_{n+1}, \dots, t_{n+m}$  to czas odpowiadający okresom wcześniejszym niż  $t_n$ , natomiast  $K_0$  odpowiada za stały poziom koncentracji, względem którego odbywają się fluktuacje. Dane  $K(t)$  dla wszystkich jonów przedstawiono na rysunku 18 - wstępna analiza tego rysunku sugeruje, iż mogą one posiadać człon stochastyczny oraz człon deterministyczny. Na rysunku można również zauważyć, że dane występują w postaci następujących po sobie wyższych oraz niższych koncentracji, dlatego aby badać trend liniowy za pomocą DFA należy zbudować szereg skumulowany. W takim razie, aby zastosować metodę DFA do analizy danych należy przeprowadzić transformację podobną do tej, ujętej we wzorach (32) oraz (33). W tym celu zdefiniowano następujący szereg skumulowany  $K_{cm}$ :

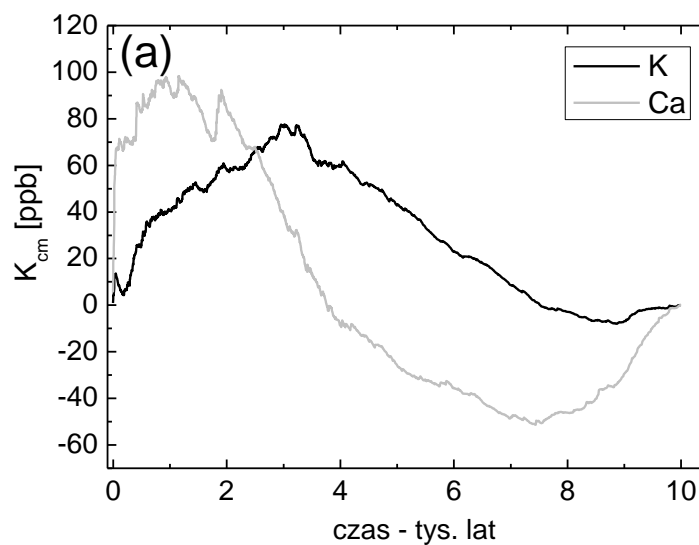
$$K_{cm}(t_k) = \sum_{i=1}^k (K(t_i) - \bar{K}), \quad (104)$$

gdzie  $\bar{K}$  to globalna średnia arytmetyczna sygnału, natomiast wielkości  $K(t_i)$  odpowiadają koncentracjom, które występowały od teraźniejszości do czasu  $t_k$ . Na rysunku 19 przedstawiono sygnał skumulowany dla sodu i magnezu -  $K_{cm}^{Na}(t), K_{cm}^{Mg}(t)$  - jonów, które, jak zasugerowano w poprzednim rozdziale, w głównej mierze pochodzą z morza. Sygnały skumulowane wyznaczono za pomocą skryptu napisanego w programie Python (dodatek A).



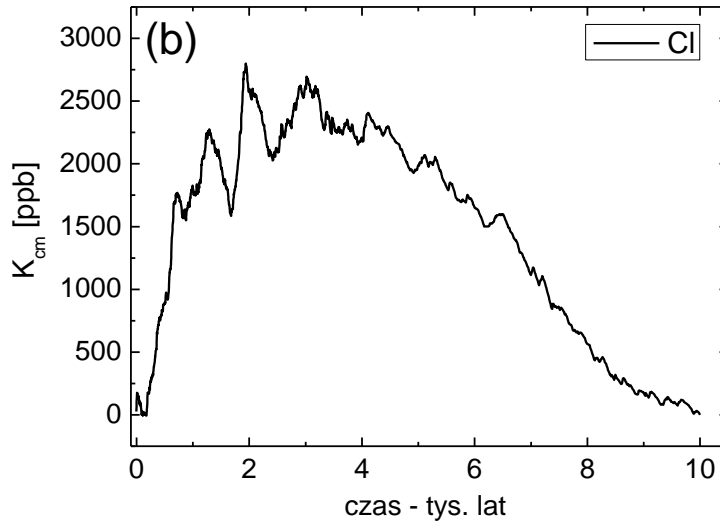
Rysunek 19. Szeregi  $K_{cm}^{Na}(t)$  oraz  $K_{cm}^{Mg}(t)$  - ostatnie 10 tys. lat, czas historyczny.

Na rysunku 20 przedstawiono sygnał skumulowany dla pozostałych jonów, które według sugestii z poprzedniego rozdziału w głównej mierze pochodzą z lądu.



Rysunek 20. Szeregi: (a) -  $K_{cm}^K(t), K_{cm}^{Ca}(t)$  - ostatnie 10 tys. lat, czas historyczny.





Rysunek 20 kont. Szereg: (b) -  $K_{cm}^{Cl}(t)$  - ostatnie 10 tys. lat, czas historyczny.

Dla jonów pochodzenia morskiego ( $K_{cm}^{Na}(t)$  oraz  $K_{cm}^{Mg}(t)$ ) widać wyraźny wzrost wartości szeregu skumulowanego za ostatnie 1.3 tys. lat, co jest związane z wyższymi koncentracjami jonów.

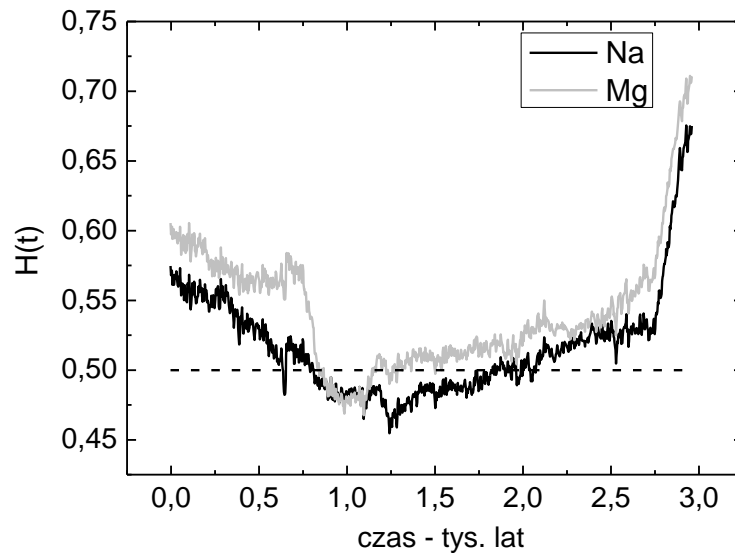
Następnie do skumulowanych szeregów czasowych zastosowano standardową procedurę DFA - wzór (24) i (25), z tą różnicą, że wartości szeregów czasowych nie były równomiernie rozmieszczone w czasie, ale odpowiadały punktom czasowym  $t_1, \dots, t_n$ , w których odczytano pomiary. Zostało to uwzględnione we wzorze (105) analogicznym do wzoru (24), z tą różnicą, że przy wyznaczaniu wartości przewidywanej przez trend liniowy użyto  $r(t_i, \tau)$ , a  $t_i$  odpowiada punktom czasowym w którym zarejestrowano dane:

$$F(\tau)^2 = \frac{\sum_{t=k}^{\tau+k-1} (x(t) - r(t, \tau))^2}{\tau}. \quad (105)$$

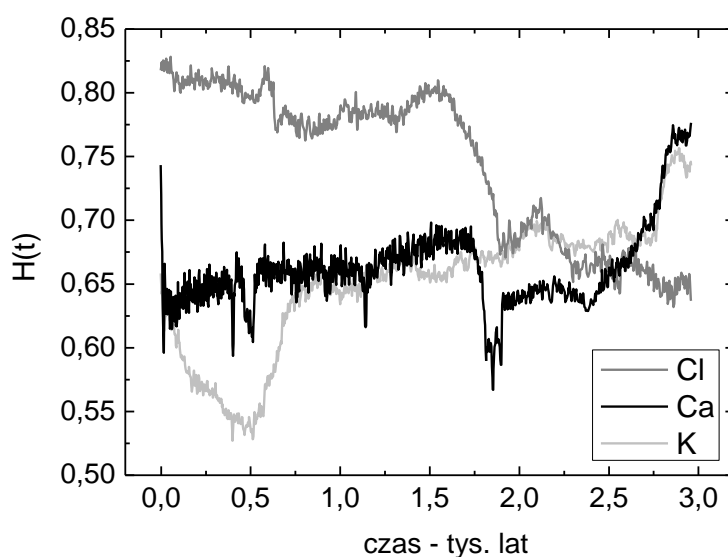
Wykładnik Hursta został wyznaczony z użyciem relacji ujętej we wzorze (28), badając wykładnik z jakim skaluje się pierwiastek ze średniego kwadratowego odchylenia w funkcji czasu. Badania przeprowadzono lokalnie dla okna o długości  $N = 1000$  pkt. Oknu temu odpowiadały różne długości czasu ze względu na różną gęstość danych, która była mniejsza dla danych starszych. Okno o długości  $N$  dzielono na mniejsze o długości  $\tau$  według następującego wzoru:

$$\tau = \left\lfloor \frac{N}{k} \right\rfloor, \quad (106)$$

gdzie  $k = 1, 2, \dots, 100$ . Dla danego punktu czasowego  $t_n$  użyto jedynie danych za okres wcześniejszy  $t_m$  taki, że  $m \geq n$ . Dlatego wartości wykładnika Hursta  $H(t_n)$  można użyć podczas próby przewidywania wartości szeregu czasowego dla okresu  $t_{n'}$  późniejszego w stosunku do  $t_n$  - dla  $n' < n$  - w analogii do wykorzystania  $H$  podczas przewidywania dynamiki danych finansowych. Ze względu na użyte okno o długości  $N$  wartości wykładnika Hursta zarejestrowano dla ostatnich 3 tysięcy lat. Wyniki obliczeń przedstawiono na rysunku 21 dla jonów o dominującym pochodzeniu morskim. Na rysunku 22 przedstawiono wyniki dla jonów o dominującym pochodzeniu lądowym. Wartości DFA wyznaczono za pomocą skryptu napisanego w programie Python (dodatek B).



Rysunek 21. Wykładnik Hursta  $H(t)$  wyliczony dla jonów o sugerowanym pochodzeniu morskim, czas historyczny w odniesieniu do teraźniejszości.



Rysunek 22. Wykładnik Hursta wyliczony dla pozostałych jonów, czas historyczny w odniesieniu do teraźniejszości.

Z analizy rysunków 21 oraz 22 można wyciągnąć następujące wnioski. Wartości wykładnika Hursta dla jonów pochodzenia morskiego oscylują wokół wartości  $H = 0.5$  oznaczającej brak auto-korelacji. Z drugiej strony cechują się one zmiennością w tym sensie, że występują okresy, dla których  $H < 0.5$  oraz okresy dla których  $H > 0.5$ . Jest to zjawisko podobne do zachowania się wykładników Hursta wyliczonych dla danych finansowych. Analizując rysunek 21 można stwierdzić, że dla okresu ok. 0.8 – 1.5 tys. lat przed chwilą obecną stwierdzono występowanie fazy anty-korelacji z  $H < 0.5$ . Począwszy od tego okresu nastąpiła stosunkowo wysoka koncentracja jonów  $\text{Na}^+$  oraz  $\text{Mg}^{2+}$ . Co jest zobrazowane przez szeregi skumulowane na rysunku 19 - sygnał anty-korelacji poprzedził wysokie wartości koncentracji jonów. W tym miejscu pojawiła się analogia do danych finansowych, dla których sygnał anty-korelacji poprzedza wysoki spadek kursów akcji. Ponadto stwierdzono duże podobieństwo w dynamice wykładnika Hursta dla sodu i magnezu, co świadczy o podobnej dynamice czynników lokalnych kształtujących obiegi tych jonów, oraz potwierdza hipotezę, iż jony magnezu w głównej mierze pochodzą tak samo jak jony sodu z soli morskiej. Najważniejszy czynnik lokalny - intensywność wiatrów wiejących nad morzem - jest w głównej mierze czynnikiem stochastycznym posiadającym niewielkie, zmienne auto-korelacje. Można tutaj doszukać się analogii do czynnika lokalnego kształtującego poszczególne kursy akcji, którym są indywidualne strategie inwestorów.

Dla pozostałych jonów, pochodzących głównie z lądu, stwierdzono inną dynamikę wykładnika Hursta. Stwierdzono występowanie dodatnich auto-korelacji zobrazowanych przez  $H > 0.5$ .

Przypuszczalnie auto-korelacja ta wynika z czynników lokalnych związanych z akumulacją pierwiastków na lądzie, co może mieć wieloletni wpływ na ich koncentrację. Jednak w przypadku lądowym, nie stwierdzono takiego podobieństwa w dynamice wykładnika Hursta, jak w przypadku morskim. Może to świadczyć o różniących się lokalnych czynnikach kształtujących koncentrację jonów pochodzenia lądowego [67, 77], np. wiatrach wiejących nad obszarami, gdzie znajdują się ropy lub osady szelfów kontynentalnych. Dokładność analizy była stosunkowo dobra, a niepewność wyznaczenia wykładnika Hursta nie przekroczyła wielkości  $\delta H = 0.0165$ , lub 2.2% otrzymanej wartości  $H$ .

Badaniom poddano również auto-korelacje dla szeregów czasowych koncentracji  $^{18}\text{O}$  oraz  $^2\text{D}$  za ostatnie 10 tys. lat. Badano dane o tej samej rozdzielczości w oknie  $N = 160$ , przyjęto podział okna według wzoru (106) dla  $k = 1, 2, \dots, 20$ . Okazało się, że występuje tu w przybliżeniu stały sygnał auto-korelacji z  $H \approx 0.75$ . Wyniku takiego można się było spodziewać, ponieważ koncentracja izotopów zależy od temperatury na Ziemi, która wykazuje auto-korelację. Badanie auto-korelacji pozwoliło na pewną ocenę dynamiki danych i zostanie ono rozszerzone w następnym rozdziale o badanie korelacji.

### 5.1.3 Analiza danych z wykorzystaniem funkcji Kopuły

W poprzednim rozdziale została omówiona analiza auto-korelacji szeregów czasowych. Obecnie zostanie pokazany dalszy ciąg analizy - badanie korelacji pomiędzy wybranymi szeregami czasowymi z użyciem funkcji Kopuły. W pierwszej kolejności przeprowadzono analizę danych silnie skorelowanych - koncentracji izotopów  $^2\text{D}$  oraz  $^{18}\text{O}$  za ostatnie 10 tys. lat. Koncentracja tych izotopów jest funkcją liniową temperatury, a ponadto ze wzorów (101) oraz (102) można wyprowadzić liniową zależność pomiędzy koncentracją  $^2\text{D}$  oraz  $^{18}\text{O}$ , co odzwierciedla ich silną korelację. Ponieważ do analizy zostały wykorzystane funkcje Kopuły, dlatego w pierwszej kolejności należało dopasować odpowiednie jednowymiarowe rozkłady brzegowe. Zaproponowano tutaj rozkład Gaussa (RG), który jest naturalnym wyborem do analizy danych stochastycznych (nawet silnie skorelowanych) oraz rozkład Johnsona  $S_U$  ( $RJS_U$ ) - wzór (97) - ponieważ dla badanych danych zarejestrowano auto-korelację, co sugeruje odejście od rozkładu Gaussa poprzez odejście od Centralnego Twierdzenia Granicznego. W tabeli 10 zaprezentowano dopasowanie oraz parametry rozkładów prawdopodobieństwa. Jednowymiarowe rozkłady prawdopodobieństwa zostały dopasowane z użyciem programu napisanego przez autora w języku Python z wykorzystaniem biblioteki Scipy (dodatek C). Aby oszacować jakość dopasowania zastosowano test Kołmogorowa - Smirnowa, w którym wartość  $d$  (supremum różnicy pomiędzy danymi oraz badanym rozkładem

prawdopodobieństwa) została użyta do wyznaczenia wartości  $p$  - prawdopodobieństwa, że dane są opisane przez badany rozkład.

Tabela 10. Parametry dopasowania jednowymiarowych rozkładów prawdopodobieństwa dla koncentracji  $^{18}\text{O}$  oraz  $^2\text{D}$ .

Izotopy	Rozkład	Wartość $p$	Wartość $d$	Parametry
$^{18}\text{O}$	RG	0.521	0.0456	$-39.350, 1.125 \ (\mu, a)$
	RJS <sub>U</sub>	<b>0.948</b>	0.0293	$12.246, 5.645, -33.177, 1.407 \ (\zeta, \nu, \mu, a)$
$^2\text{D}$	RG	0.460	0.0476	$-310.05, 8.876$
	RJS <sub>U</sub>	<b>0.845</b>	0.0345	$13.618, 6.676, -252.801, 14.976$

Na poziomie istotności 0.05 nie można odrzucić żadnego rozkładu, jednakże do dalszej analizy danych wybrano rozkład RJS<sub>U</sub> posiadający wyższe wartości  $p$ . Odejście od rozkładu Gaussa można uzasadnić auto-korelacją danych - rozdział 4.2. Silna korelacja pomiędzy danymi odzwierciedla wysoką wartość współczynnika korelacji  $\tau$  Kendall'a równą  $\tau = 0.910$ . Do danych dopasowano funkcje Kopuły omówione w rozdziale 3.2, zgodnie z procedurą korzystającą ze współczynnika korelacji  $\tau$  Kendall'a - procedury omówionej w rozdziale 3.4.3 - wzór (68, 72). Dopasowanie Kopuły Studenta przeprowadzono według procedury opisanej w rozdziale 3.4.4, z wykorzystaniem wzoru (84). W przypadku Kopuły Studenta procedura dopasowania wskazała na Kopułę Gaussowską, otrzymaną z Kopuły Studenta dla  $\nu \rightarrow \infty$ . W celu dopasowania oraz generowania danych przy pomocy Kopuł Archimedesowskich (Claytona, Gumbela, Franka) oraz Kopuł odwróconych (odwrócona Claytona i odwrócona Gumbela) zastosowano skrypt napisany w programie Python (dodatek D) wzorując się na ogólnie dostępnej bibliotece copula.py. Kopuły odwrócone zdefiniowano przy pomocy transformacji ujętej we wzorze (79). Do dopasowania oraz generowania danych przy pomocy Kopuły Studenta zastosowano skrypt napisany w programie Python (dodatek E).

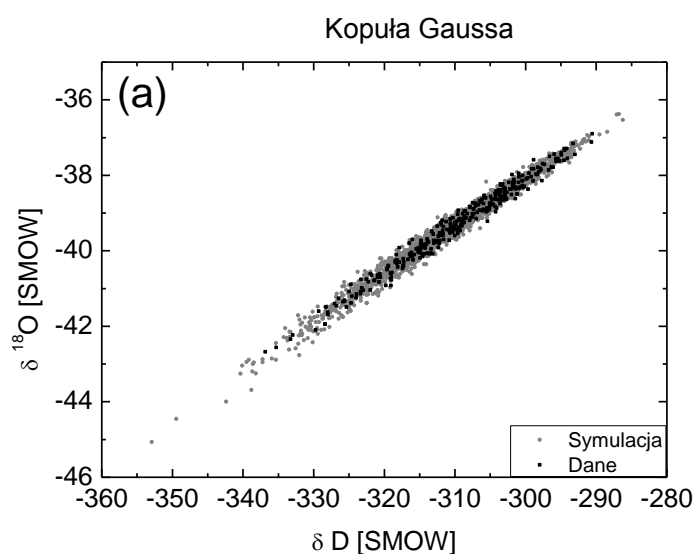
W celu oceny jakości dopasowania poszczególnych Kopuł, zbadano w jaki sposób opisują one ekstremalne dane o bardzo niskich lub bardzo wysokich koncentracjach  $^{18}\text{O}$  oraz  $^2\text{D}$ . Dlatego zbadano indeks dolnej oraz górnej zależności ogonowej. Do wyliczenia indeksów dla poszczególnych Kopuł zastosowano wzory ujęte w tabelach 3, 4 oraz 5. Wartości zależności ogonowych porównano z wartościami empirycznymi, wyznaczonymi z badanych szeregów zgodnie ze wzorami (75) i (78), przy założeniu progu  $\alpha = 0.025$ . W tabeli 11 pokazano indeksy zależności ogonowej, zarówno empiryczne jak i indeksy wyliczone dla dopasowanych funkcji Kopuły. Empiryczne wartości indeksów

wyliczono z użyciem skryptu napisanego w programie Python (dodatek F). Warto przypomnieć, że Kopuła Gaussa oraz Franka nie posiadają indeksów zależności ogonowej.

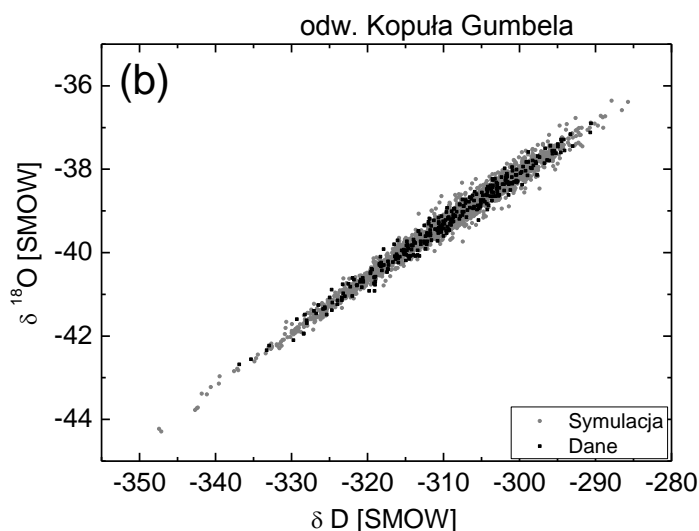
Tabela 11. Indeks zależności ogonowej dla  $^{18}\text{O}$  oraz  $^2\text{D}$ .

Zależność	Empiryczne	Kopuła			
		Claytona	Gumbela	odw. Claytona	odw. Gumbela
$\lambda_u$	0.51	-	0.94	0.97	-
$\lambda_l$	0.63	0.97	-	-	0.94

W tym miejscu należy zauważyć, że żadna z funkcji Kopuły nie modeluje właściwie indeksów zależności ogonowej. Dane wygenerowane przez funkcje Kopuły oraz dane empiryczne poddano ocenie wizualnej; wybrano Kopułę Gaussowską z parametrem  $\rho = 0.990$  i odwróconą Kopułę Gumbela z parametrem  $\theta = 11.17$  - rysunek 23.



Rysunek 23. Dopasowanie funkcji Kopuły do koncentracji izotopu tlenu ( $^{18}\text{O}$ ) i deuteru ( $^2\text{D}$ ) - (a) - Kopuła Gaussowska.



Rysunek 23 kont. Dopasowanie funkcji Kopuły do koncentracji izotopu tlenu ( $^{18}\text{O}$ ) i deuteru ( $^2\text{D}$ ) - (b) - odwrócona Kopuła Gumbela.

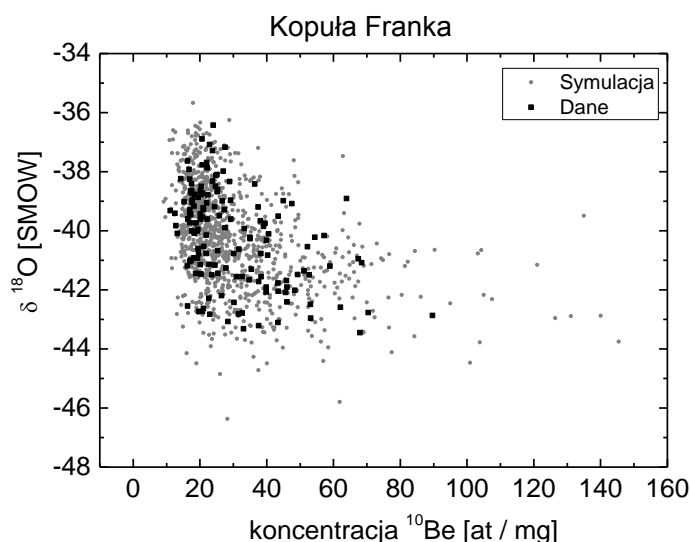
Jak już wspomniano wcześniej nie udało się znaleźć funkcji Kopuły, która dobrze modeluje dane ekstremalne. Trudność w znalezieniu właściwej funkcji Kopuły wynika z faktu, że koncentracje tlenu  $^{18}\text{O}$  oraz deuteru są ze sobą powiązane w sposób praktycznie liniowy - deterministyczny. Czynniki globalny związany z temperaturą na Ziemi ma tutaj dużą przewagę nad czynnikami lokalnymi, powodującymi stochastyczne zachowanie się zmiennych. Zależność liniowa pomiędzy koncentracjami wprowadza indeksy zależności ogonowej, które nie mogą być modelowane przez Kopułę Gaussowską dobrze opisującą pozostałe dane. Natomiast stwierdzone auto-korelacje mogą być wykorzystane w procedurze wyboru właściwej funkcji Kopuły. Globalna auto-korelacja pokazana w poprzednim rozdziale świadczy o tworzeniu się klasterów danych ekstremalnych [64] - po dużych koncentracjach bardziej prawdopodobne są również duże koncentracje i vice versa. Dlatego po zarejestrowaniu małych koncentracji - związanych z niższą temperaturą na Ziemi - wystąpienie kolejnych małych koncentracji jest bardziej prawdopodobne, co sugeruje zastosowanie Kopuły posiadającej wysoki indeks dolnej zależności ogonowej - np. odwróconej Kopuły Gumbela.

Przeanalizujmy teraz całkowicie odmienny przypadek, koncentrację  $^{18}\text{O}$  oraz  $^{10}\text{Be}$  za ostatnie 225 tys. lat. Do danych dopasowano jednowymiarowe rozkłady brzegowe, tak samo jak w poprzednim przypadku, gdy zaproponowano (RG) oraz (RJS<sub>U</sub>) - tabela 12.

Tabela 12. Parametry dopasowania jednowymiarowych rozkładów prawdopodobieństwa dla koncentracji  $^{18}\text{O}$  oraz  $^{10}\text{Be}$ .

Pierwiastki	Rozkład	Wartość p	Parametry
$^{18}\text{O}$	RG	<b>0.551</b>	$-40.247, 1.604 (\mu, \sigma)$
	RJS <sub>U</sub>	<b>0.551</b>	$-61 \cdot 10^3, 33 \cdot 10^5, -97 \cdot 10^3, 52 \cdot 10^5 (\zeta, \nu, \mu, \sigma)$
$^{10}\text{Be}$	RG	0.00036	29.829, 14.303
	RJS <sub>U</sub>	<b>0.432</b>	-1.912, 1.114, 14.368, 3.954

W przypadku  $^{10}\text{Be}$  rozkład Johnsona  $S_U$  przedstawia dobry model danych, w przypadku  $^{18}\text{O}$  rozkład Johnsona  $S_U$  modeluje dane równie dobrze jak rozkład Gaussa. Jednakże, aby zastosować ten sam rozkład brzegowy i utworzyć jednolity model wybrano rozkład Johnsona  $S_U$ . Co ważne, rozkładu tego nie można odrzucić na poziomie istotności 0.05. Następnie rozpoczęto analizę danych dwuwymiarowych z zastosowaniem funkcji Kopuły. Stwierdzono słabą, ujemną korelację pomiędzy danymi,  $\tau = -0,277$ . Do danych dopasowano funkcje Kopuły omówione w rozdziale 3.2, według procedury z użyciem współczynnika korelacji  $\tau$  Kendall'a omówionej w podrozdziale 3.4.4. Do analizy danych o ujemnej korelacji w miejsce Kopuły Gumbela zastosowano przekręconą Kopułę Gumbela zdefiniowaną transformacją (82). Ze względu na ujemną korelację nie przeprowadzono badań uwzględniających indeksy zależności ogonowych. Według oceny wizualnej najlepszy model dla danych reprezentuje Kopuła Franka z parametrem  $\theta = -2.659$  - rysunek 24.



Rysunek 24. Dopasowanie funkcji Kopuły do koncentracji izotopu tlenu ( $^{18}\text{O}$ ) i berylu ( $^{10}\text{Be}$ ).



Niewielka ujemna korelacja pomiędzy danymi świadczy o tym, iż czynniki deterministyczne kształtujące koncentrację berylu  $^{10}\text{Be}$  - aktywność Słońca, natężenie promieniowania kosmicznego oraz natężenie pola magnetycznego Ziemi - są zbyt słabo związane z temperaturą na Ziemi, czyli czynnikiem deterministycznym kształtującym koncentrację tlenu  $^{18}\text{O}$ , aby przeważać wpływ czynników stochastycznych [67, 76].

Następnie przeprowadzono analizę koncentracji jonów  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ . Do szeregów czasowych opisujących koncentrację jonów dopasowano, tak jak poprzednio, rozkłady prawdopodobieństwa (RG) oraz ( $\text{RJS}_U$ ). Parametry te przedstawiono w tabeli 13. Za zastosowaniem innego rozkładu niż RG przemawiają zmienne i niezerowe auto-korelacje.

Tabela 13. Parametry dopasowania jednowymiarowych rozkładów prawdopodobieństwa dla koncentracji  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ .

Jony	Rozkład	Wartość $p$	Wartość $d$	Parametry
$\text{Na}^+$	RG	0	0.117	10.084, 4.190 ( $\mu, a$ )
	$\text{RJS}_U$	<b>0.628</b>	0.0165	-1.585, 1.562, 5.424, 3.163 ( $\zeta, v, \mu, a$ )
$\text{K}^+$	RG	0	0.133	0.764, 0.323
	$\text{RJS}_U$	<b>0.576</b>	0.0171	-1.739, 1.394, 0.426, 0.163
$\text{Mg}^{2+}$	RG	0	0.102	1.529, 0.562
	$\text{RJS}_U$	<b>0.636</b>	0.0163	-1.585, 1.693, 0.878, 0.505
$\text{Cl}^-$	RG	$6.88 \cdot 10^{-15}$	0.0894	38.951, 12.377
	$\text{RJS}_U$	<b>0.992</b>	0.0095	-2.307, 1.991, 19.758, 11.770
$\text{Ca}^{2+}$	RG	0	0.221	0.984, 0.692
	$\text{RJS}_U$	0.029	0.0319	-1.109, 1.086, 0.624, 0.184

Analizując tabelę 13, można stwierdzić, iż dla wszystkich jonów, z wyjątkiem wapnia, rozkłady brzegowe są dobrze opisane przez rozkład Johnsona  $S_U$ . W przypadku wapnia (według testu Kołmogorowa - Smirnowa) prawdopodobieństwo, że dopasowany rozkład  $\text{RJS}_U$  opisuje dane wynosi 0.029, co jest podstawą odrzucenia tego rozkładu na poziomie istotności równym 0.05. Dlatego w przypadku  $\text{Ca}^{2+}$ , w dalszej analizie, zostanie zastosowana dystrybucja empiryczna. Dane koncentracji wapnia -  $K^{Ca}$  zostaną uszeregowane rosnąco, oraz zakładając, że każda wartość jest równo prawdopodobna, każdej wartości  $K^{Ca}$  zostanie przyporządkowane prawdopodobieństwo, iż zmienna losowa  $x$  jest mniejsza lub równa od niej, co można zapisać poprzez:

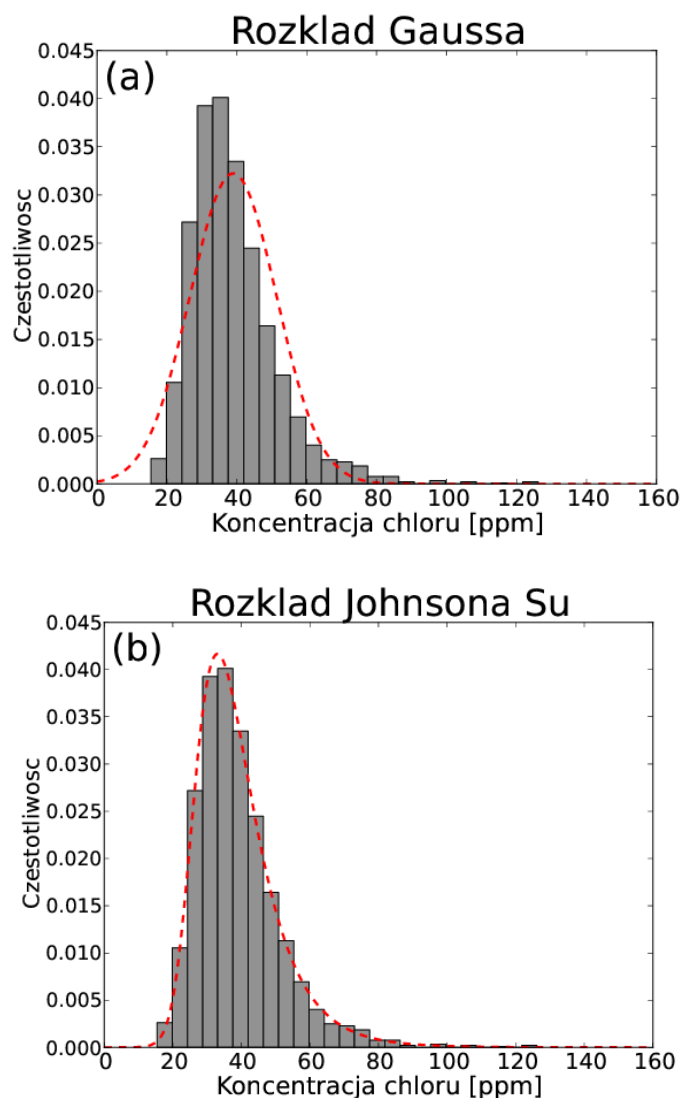
$$F(K^{Ca}) = P(x \leq K^{Ca}). \quad (107)$$

Pozwoli to przetransformować dane do przestrzeni  $[0, 1]$  w celu analizy z użyciem funkcji Kopuły. Możliwa jest również odwrotna transformacja danych

$$K^{Ca} = F^{-1}(P(x \leq K^{Ca})), \quad (108)$$

umożliwiająca transformację danych wygenerowanych przez funkcje Kopuły do przestrzeni koncentracji jonów wapnia. Różnica w rozkładach prawdopodobieństwa pomiędzy wapniem i innymi jonami wynika z faktu występowania w rozkładzie prawdopodobieństwa  $Ca^{2+}$  dużych wartości znacznie odbiegających od większości danych opisujących koncentrację  $Ca^{2+}$ . Może być to związane z faktem, że jony wapnia są głównie pochodzenia lądowego z osadów szelfów kontynentalnych [67, 77], a mechanizm obiegu wapnia może powodować sporadyczne występowanie dużych fluktuacji koncentracji. Na rysunku 25 przedstawiono przykład dopasowania rozkładu Gaussa oraz Johnsona  $S_U$  do danych koncentracji chloru. Z rysunku widać, że rozkład Johnsona  $S_U$  znacznie lepiej modeluje dane, ponieważ jest niesymetryczny.

Po przetransformowaniu do przestrzeni  $[0, 1]^2$  utworzono pary reprezentujące zróżnicowane podobieństwo danych. Na początku wszystkie dane porównywano z koncentracją jonów sodu, które są pochodzenia morskiego. Najbardziej podobne dane powinny odpowiadać koncentracji magnezu, co potwierdza hipotezę, że jony magnezu są w większości pochodzenia morskiego a dynamika ich obiegu jest bardzo podobna do dynamiki obiegu jonów sodu [67] - występuje podobieństwo w czynnikach lokalnych. Zmienna dwuwymiarowa  $[K^{Na}, K^{Mg}]$  jest szczególnie interesująca, ponieważ znaleziono analogię pomiędzy dynamiką danych finansowych oraz obiegiem sodu i magnezu. Najmniej podobne dane powinna stanowić para  $[K^{Na}, K^{Ca}]$ , ponieważ występuje podstawowa różnica w pochodzeniu tych jonów - jony wapnia są pochodzenia lądowego. Pewna korelacja pomiędzy koncentracją sodu i wapnia może być związana z występowaniem czynnika globalnego. Pozostałe pary  $[K^{Na}, K^K]$  oraz  $[K^{Na}, K^{Cl}]$  powinny odpowiadać pośredniemu przypadkowi, ponieważ założono, że jony potasu oraz chloru, mimo, że są głównie pochodzenia lądowego, w pewnej części pochodzą z morza. Następnie wyliczono współczynnik korelacji  $\tau$  Kendalla pomiędzy koncentracją jonów sodu oraz koncentracją pozostałych jonów. Wyniki przedstawiono w tabeli 14 w kolejności malejącej korelacji.



Rysunek 25. Dopasowanie rozkładu (a) Gaussa oraz (b) Johnsona  $S_U$  do danych obrazujących koncentrację chloru.

Tabela 14. Współczynnik korelacji  $\tau$  Kendall'a pomiędzy koncentracją jonów sodu oraz pozostałych jonów.

Jony	$Mg^{2+}$	$K^+$	$Cl^-$	$Ca^{2+}$
$\tau$ Kendall'a	0.790	0.490	0.408	0.402

Tak jak można się było tego spodziewać, widać tutaj silniejszą korelację między koncentracją jonów  $Na^+$  i  $Mg^{2+}$  oraz słabszą pomiędzy koncentracją jonów  $Na^+$  i  $Ca^{2+}$ . Niezerowa korelacja pomiędzy koncentracją sodu i wapnia świadczy o występowaniu czynnika globalnego wpływającego na koncentrację zarówno jonów pochodzenia lądowego jak i jonów pochodzenia morskiego, który

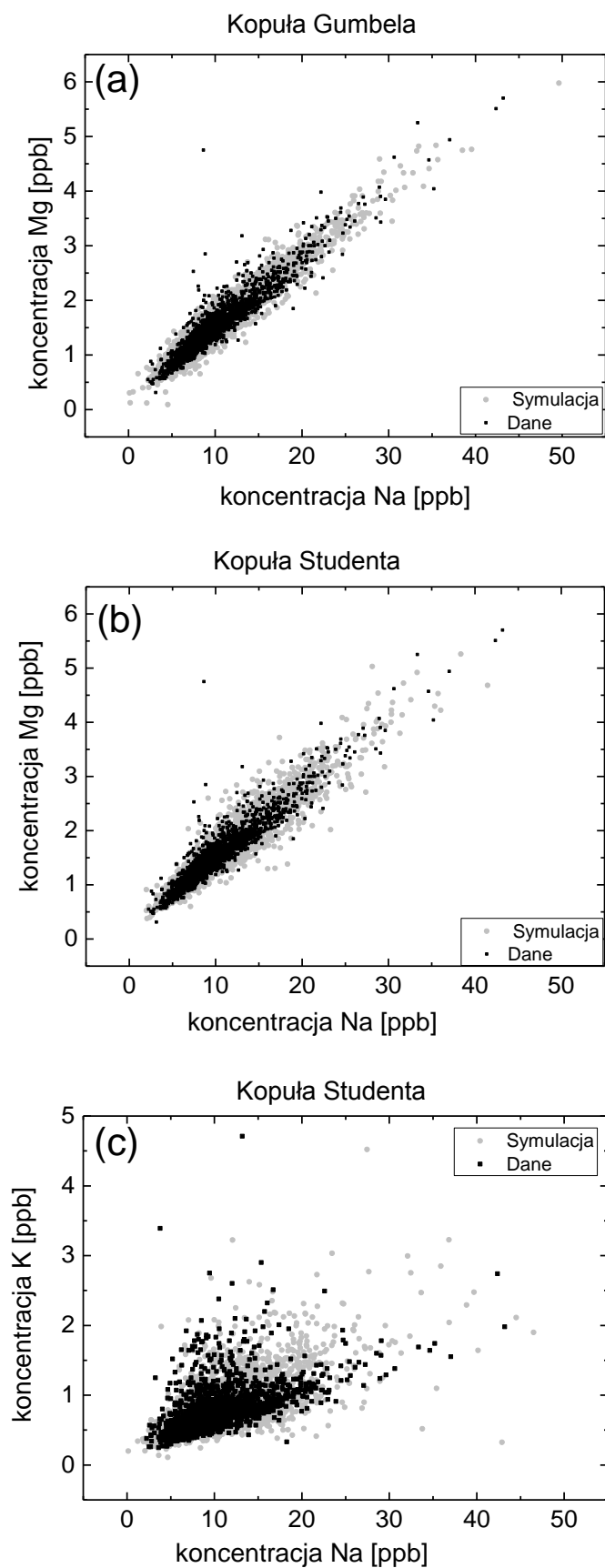
jest związany z występowaniem silnych układów cyklonicznych, odpowiedzialnych za transport aerozoli w kierunku południowym. Pośredni przypadek jonów  $\text{Na}^+$  oraz  $\text{K}^+$  świadczy o tym, że mimo iż jony potasu pochodzą przede wszystkim z lądu, to ich część pochodzi z morza. Niestety, trudno wnioskować na podstawie tabeli 14 o morskim pochodzeniu chloru, gdyż korelacja pomiędzy koncentracją sodu i chloru jest podobna do tej pomiędzy koncentracją sodu i wapnia.

Do danych tak jak poprzednio dopasowano funkcje Kopuły (rysunek 26) oraz porównano indeksy zależności ogonowych (tabela 15).

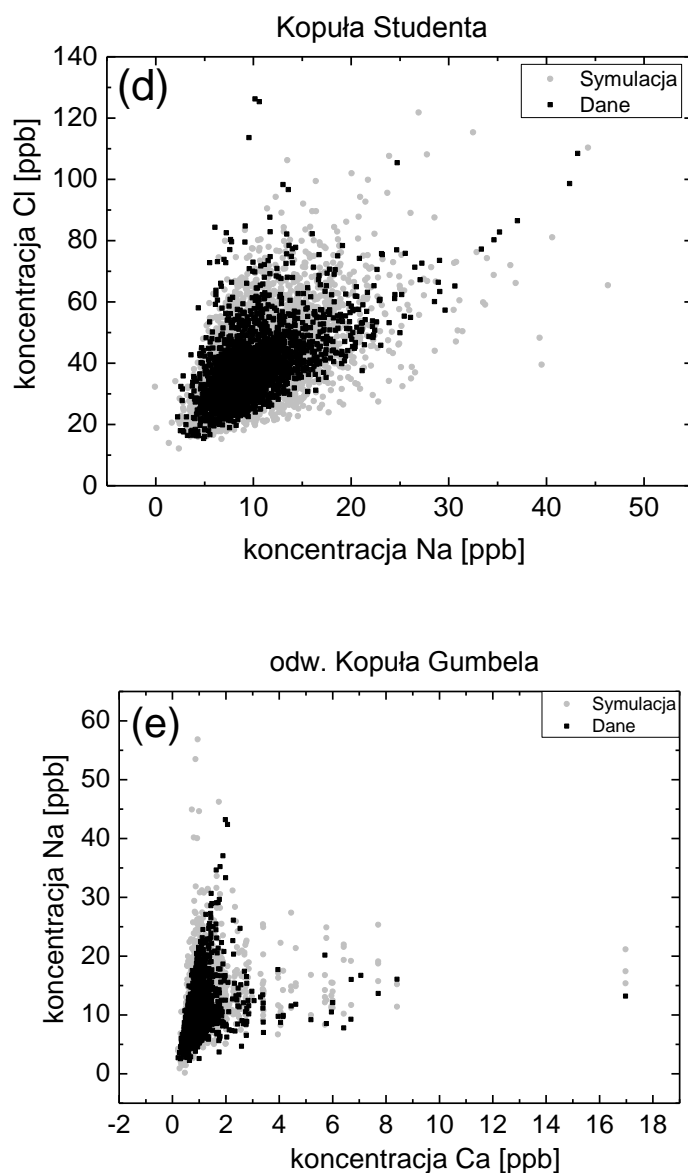
Tabela 15. Indeksy zależności ogonowych dla poszczególnych funkcji Kopuły.

Pary jonów	Zależność ogonowa	Empiryczna	Kopuła		
			Gumbela	odw. Gumbela	Studenta
$\text{Na}^+, \text{Mg}^{2+}$	$\lambda_u$	0.81	0.843	-	0.695
	$\lambda_l$	0.75	-	0.843	0.695
$\text{Na}^+, \text{K}^+$	$\lambda_u$	0.21	0.575	-	0.368
	$\lambda_l$	0.44	-	0.575	0.368
$\text{Na}^+, \text{Cl}^-$	$\lambda_u$	0.31	0.492	-	0.130
	$\lambda_l$	0.29	-	0.492	0.130
$\text{Na}^+, \text{Ca}^{2+}$	$\lambda_u$	0.02	0.486	-	0.423
	$\lambda_l$	0.31	-	0.486	0.423

W przypadku pary sodu i magnezu Kopuła Gumbela dobrze modeluje górną zależność ogonową. Warto w tym miejscu zauważyć, że duże koncentracje tych jonów wystąpiły po sygnale anty-korelacji (w okresie ostatnich 1.3 tys. lat). W celu lepszej analizy wysokich koncentracji warto zbadać rozkład wartości ekstremalnych dla pary  $[K^{Na}, K^{Mg}]$ . Z drugiej strony, jeżeli poddać analizie wszystkie dane oraz oba indeksy zależności ogonowej, możemy stwierdzić, że ogólnie Kopuła Studenta najlepiej modeluje dane sodu i magnezu. W przypadku pary  $[K^{Na}, K^K]$  oraz  $[K^{Na}, K^{Cl}]$ , Kopuła Studenta najlepiej modeluje dane skrajne, natomiast w przypadku pary  $[K^{Na}, K^{Ca}]$  najlepsza jest odwrócona Kopuła Gumbela.



Rysunek 26. Koncentracje jonów oraz dane wygenerowane przez funkcje Kopuły: (a, b) -  $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ;  
(c) -  $\text{Na}^+$ ,  $\text{K}^+$ .



Rysunek 26 kont. Koncentracje jonów oraz dane wygenerowane przez funkcje Kopuły: (d) -  $\text{Na}^+$ ,  $\text{Cl}^-$ ; (e) -  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ .

Analizując rysunek 26 można stwierdzić, iż proponowane funkcje Kopuły stanowią dobry model dla danych. Podobieństwo wizualne pomiędzy parami  $[K^{\text{Na}}, K^K]$  i  $[K^{\text{Na}}, K^{\text{Cl}}]$ , uwidocznione na rysunku 26, wraz z podobnym modelem funkcji Kopuły dla pary  $[K^{\text{Na}}, K^K]$  oraz  $[K^{\text{Na}}, K^{\text{Cl}}]$ , sugeruje podobieństwo czynników lokalnych dla potasu i chloru. Wynika to również z faktu, iż chlor jest w pewnej części pochodzenia morskiego. Dla pary  $[K^{\text{Na}}, K^{\text{Ca}}]$  stwierdzono brak górnej zależności ogonowej, a na wykresie 26 nie można zobaczyć korelacji dla dużych koncentracji tych jonów.

Oznacza to mały wpływ czynnika globalnego dla dużych koncentracji. Parametry funkcji Kopuły przedstawiono w tabeli 16.

Tabela 16. Parametry funkcji Kopuły

Pary jonów	Kopuła	Parametry	
$[\text{Na}^+, \text{Mg}^{2+}]$	Studenta	$\rho = 0.946$	$\nu = 5$
$[\text{Na}^+, \text{Mg}^{2+}]$	Gumbela	$\theta = 4.75$	
$[\text{Na}^+, \text{K}^+]$	Studenta	$\rho = 0.695$	$\nu = 7$
$[\text{Na}^+, \text{Cl}^-]$	Studenta	$\rho = 0.598$	$\nu = 14$
$[\text{Na}^+, \text{Ca}^{2+}]$	odw. Gumbela	$\theta = 1.67$	

Poniżej zostaną jeszcze przedstawione wyniki analizy porównawczej z koncentracją jonów wapnia, o dominującym pochodzeniu lądowym. W pierwszej kolejności zostało wykonane porównanie występowania korelacji pomiędzy koncentracjami jonów (tabela 17).

Tabela 17. Współczynnik korelacji  $\tau$  Kendall'a pomiędzy koncentracją jonów wapnia oraz pozostałych jonów.

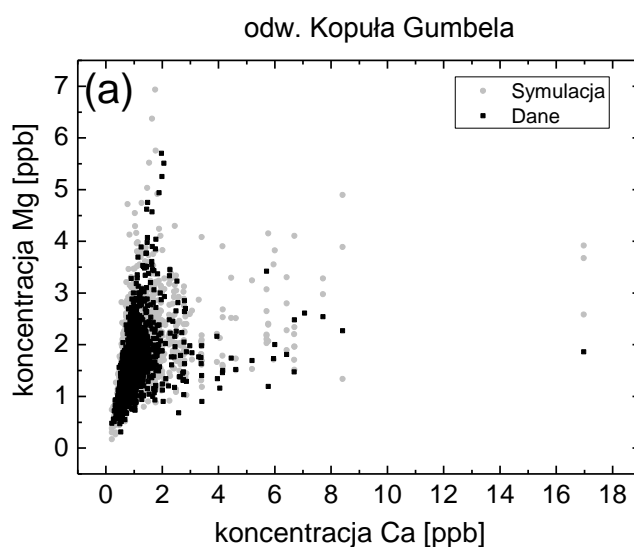
Jony	$\text{Mg}^{2+}$	$\text{K}^+$	$\text{Na}^+$	$\text{Cl}^-$
$\tau$ Kendall'a	0.478	0.415	0.402	0.230

Analizując tabelę 17 można zaobserwować ogólnie słabsze korelacje niż w przypadku tabeli 14, gdzie występowało podobieństwo czynników lokalnych, związanych z obiegiem jonów pochodzenia morskiego. Świadczy to o braku podobieństwa czynników lokalnych odpowiadających za obieg jonów pochodzenia lądowego. Podobny czynnik lokalny, który tu może występować, jest związany z jednoczesnym pochodzeniem jonów wapnia i magnezu z osadów szelfów kontynentalnych [77]. Niska korelacja pomiędzy koncentracją jonu wapnia i chloru może wynikać z całkiem innych mechanizmów obiegu tych pierwiastków. Do danych, tak jak poprzednio, dopasowano funkcje Kopuły oraz porównano indeksy zależności ogonowych.

Tabela 18. Indeksy zależności ogonowych dla poszczególnych funkcji Kopuły.

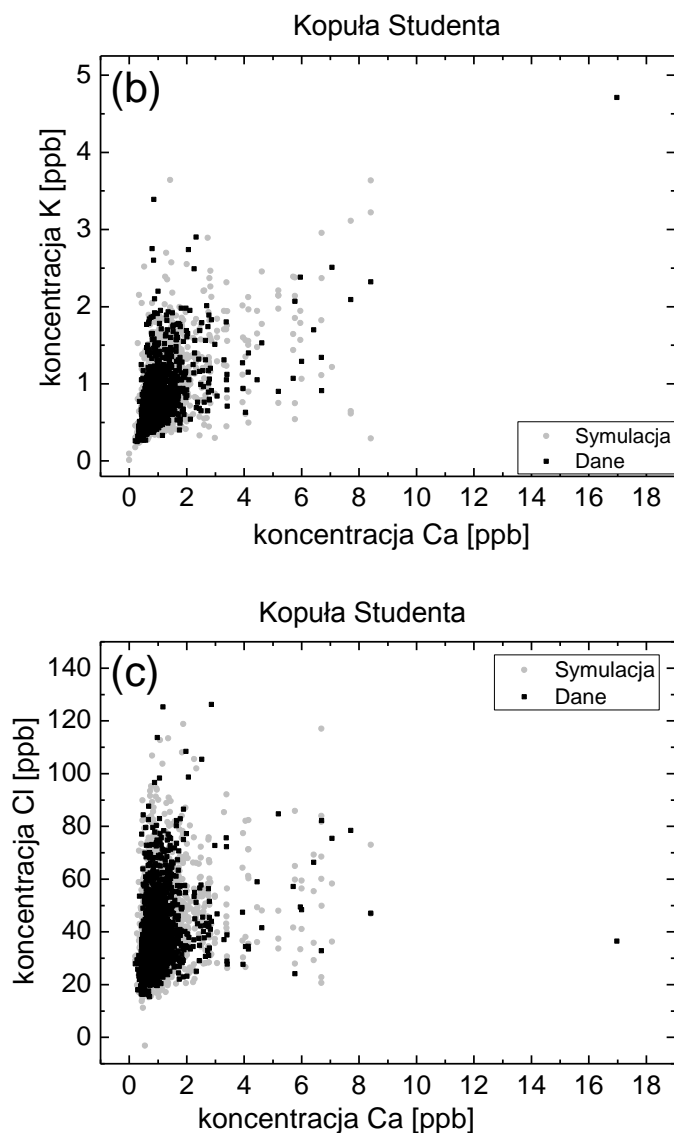
Pary jonów	Zależność ogonowa	Empiryczna	Kopuła		
			Gumbela	odw. Gumbela	Studenta
[Ca <sup>+</sup> , Mg <sup>2+</sup> ]	$\lambda_u$	0.04	0.565	-	0.483
	$\lambda_l$	0.35	-	0.565	0.483
[Ca <sup>+</sup> , K <sup>+</sup> ]	$\lambda_u$	0.25	0.500	-	0.434
	$\lambda_l$	0.35	-	0.500	0.434
[Ca <sup>+</sup> , Cl <sup>-</sup> ]	$\lambda_u$	0.17	0.294	-	0.091
	$\lambda_l$	0.19	-	0.294	0.091

Analizując tabele 18, można zauważyć brak indeksu górnej zależności ogonowej dla pary  $[K^{Ca}, K^{Mg}]$  podobnie jak dla pary wapnia i sodu (drugiego obok magnezu, jonu pochodzącego z morza). Dla pary  $[K^{Ca}, K^{Mg}]$  zależności ogonowe najlepiej modeluje odwrócona Kopuła Gumbela, natomiast dla pozostałych par Kopuła Studenta. Opisane dane dwuwymiarowe oraz modele Funkcji Kopuły przedstawiono na rysunku 27.



Rysunek 27. Koncentracje jonów oraz dane wygenerowane przez funkcje Kopuły: (a) - Mg<sup>2+</sup>, Ca<sup>2+</sup>.





Rysunek 27 kont. Koncentracje jonów oraz dane wygenerowane przez funkcje Kopuły: (b) -  $K^+$ ,  $Ca^{2+}$ ; (c) -  $Cl^-$ ,  $Ca^{2+}$ .

Analizując rysunek 27 można stwierdzić, iż proponowane funkcje Kopuły stanowią dobry model dla danych. Dla pary zawierającej wapń i magnez można stwierdzić brak korelacji dla wysokich koncentracji. Potwierdza to zaproponowaną wcześniej hipotezę, że czynnik globalny wpływający zarówno na koncentrację jonów pochodzenia morskiego jak i lądowego, nie ma istotnego wpływu przy dużych koncentracjach. Parametry funkcji Kopuły przedstawiono w tabeli 19.

Tabela 19. Parametry funkcji Kopuły

Pary jonów	Kopuła	Parametry	
$[Ca^{2+}, Mg^{2+}]$	odw. Gumbela	$\theta = 1.917$	
$[Ca^{2+}, K^+]$	Studenta	$\rho = 0.607$	$\nu = 3$
$[Ca^{2+}, Cl^-]$	Studenta	$\rho = 0.353$	$\nu = 9$

Na koniec rozdziału można podać następujące wnioski wynikające z przeprowadzonych badań:

- magnez, tak samo jak sód, jest jonem pochodzenia głównie morskiego, co wynika z podobnej dynamiki wykładnika Hursta oraz korelacji między koncentracjami tych jonów,
- korelacja dla jonów pochodzenia lądowego jest mniejsza niż dla jonów pochodzenia morskiego, co świadczy o różnych lokalnych czynnikach związanych z mechanizmem pochodzenia lądowego tych jonów,
- czynnik globalny, kształtujący zarówno koncentrację jonów pochodzenia morskiego jak i lądowego, którym jest siła układów cyklonicznych odpowiedzialnych za transport aerozoli w kierunku południowym, nie ma istotnego wpływu na dane w przypadku dużych koncentracji,
- koncentracja jonów pochodzenia lądowego wykazuje wyraźną auto-korelację, co może wynikać z akumulacji jonów na lądzie,
- koncentracja jonów pochodzenia morskiego ma zmienną auto-korelację analogiczną do szeregów czasowych danych finansowych.

W następnym podrozdziale zostanie przeprowadzona analiza danych ekstremalnych w celu uzupełnienia badań, szczególnie dla jonów pochodzenia morskiego, których dynamika wydaje się być analogiczna do dynamiki danych finansowych.

#### 5.1.4 Analiza danych ekstremalnych

Przeanalizujemy wzajemną relację koncentracji  $[K^{Na}, K^{Mg}]$  w analogii do danych finansowych. W rozdziale 5.1.2 pokazano, iż po sygnale anty-korelacji występowały wysokie wartości zarówno  $K^{Na}$  jak i  $K^{Mg}$ . Analogicznie po sygnale anty-korelacji prawdopodobne są duże spadki kursów akcji – wysokie ujemne stopy zwrotu [2]. Dlatego do dalszych badań zostaną zaproponowane pary danych maksymalnych  $[K_{eks}^{Na}(t), K_{eks}^{Mg}(t)]$  spełniające dla danego czasu  $t$  jednocześnie warunek  $K^{Na}(t) \geq 17.5$  oraz  $K^{Mg}(t) \geq 2.5$ . Do danych maksymalnych zastosowano standardową procedurę

dopasowania funkcji Kopuły. W tabeli 20 pokazano parametry jednowymiarowych rozkładów brzegowych. Do analizy danych wybrano rozkład  $RJS_U$ .

Tabela 20. Parametry dopasowania jednowymiarowych rozkładów prawdopodobieństwa.

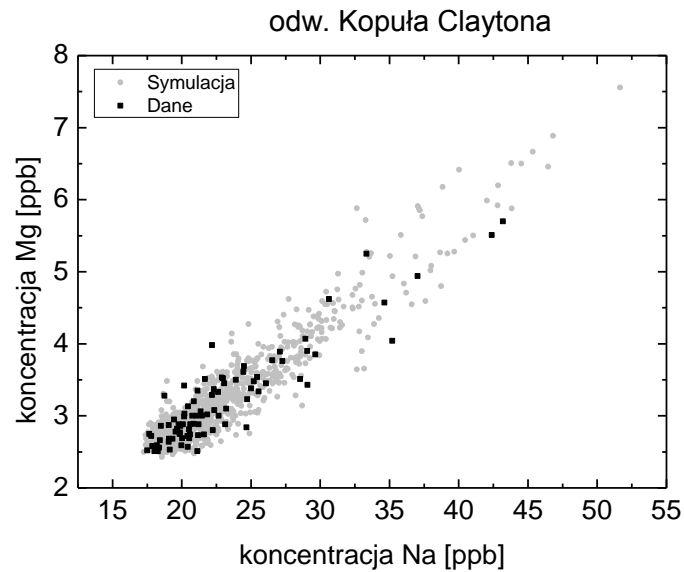
Jony	Rozkład	Wartość $p$	Wartość $d$	Parametry
$Na^+$	RWE	0	0.324	$-1, 22.886 \ 5.111(\xi, \mu, a)$
	$RJS_U$	<b>0.947</b>	0.0558	$-6.185, 1.271, 16.868, 0.0686 (\zeta, \nu, \mu, a)$
$Mg^{2+}$	RWE	0	0.266	$-1, 3.205, 0.664$
	$RJS_U$	<b>0.673</b>	0.0769	$-6.389, 1.195, 2.407, 0.00548$

W tabeli 21 przedstawiono indeks górnej zależności ogonowej dla danych empirycznych, wyznaczony według wzoru (75), dla  $\alpha = 0.025$ . Nie badano indeksu dolnej zależności ogonowej ze względu na ten fakt, że badano tu dane maksymalne.

Tabela 21. Indeks górnej zależności ogonowej dla ekstremalnych koncentracji jonów.

Para jonów	Zależność ogonowa	Empiryczna	Kopuła	
			Gumbela	odw. Claytona
$Na^+, Mg^{2+}$	$\lambda_u$	0.909	0.752	0.850

Do analizy danych zaproponowano odwróconą Kopułę Claytona, która najlepiej opisuje górną zależność ogonową.



Rysunek 28. Maksymalne koncentracje jonów oraz dane wygenerowane przez funkcję Kopuły.

Na rysunku 28 zaprezentowano dopasowanie funkcji Kopuły do maksymalnych koncentracji sodu oraz magnezu. Podczas dopasowania danych otrzymano współczynnik korelacji  $\tau = 0.681$  oraz parametr odwróconej Kopuły Claytona  $\theta = 4.263$ . Można stwierdzić wysoką korelację oraz wysoki indeks górnej zależności ogonowej. W dalszej części pracy wyniki analizy danych ekstremalnych zostaną porównane z analizą danych finansowych oraz zostaną przedyskutowane podobieństwa i różnice.

Warto w tym miejscu odnieść się do hipotezy braku korelacji dla wysokich koncentracji jonu pochodzenia morskiego oraz lądowego. Analizując dane ekstremalne koncentracji sodu i wapnia  $[K_{eks}^{Na}, K_{eks}^{Ca}]$  takie, że  $K^{Na} \geq 8$  oraz  $K^{Ca} \geq 1.8$  stwierdzono brak korelacji, ze współczynnikiem korelacji  $\tau = -0.095$ . Podobny wynik można otrzymać dla ekstremalnych koncentracji magnezu i wapnia  $[K_{eks}^{Mg}, K_{eks}^{Ca}]$ . Wyniki te potwierdzają hipotezę o braku jednoczesnego wpływu globalnego czynnika na wysokie koncentracje jonów pochodzenia morskiego i lądowego.

## 5.2. Badanie maksymalnych wezbrań sztormowych w Bałtyku

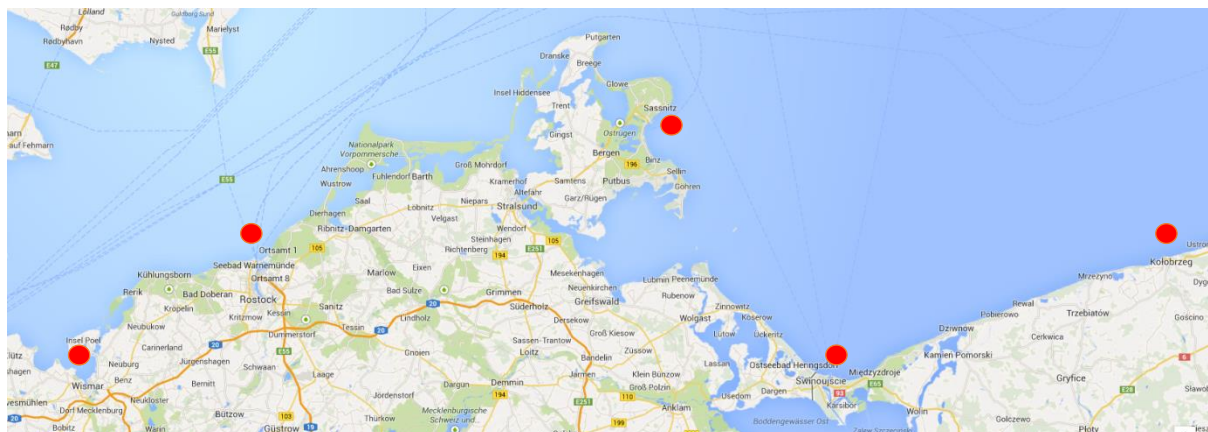
W poprzednich podrozdziałach omówiono badanie szeregów czasowych odzwierciedlających koncentrację jonów oraz izotopów w rdzeniach arktycznych. Obecnie zostanie omówiony inny rodzaj danych geofizycznych - maksymalne wezbrania sztormowe.

### 5.2.1 Czynniki naturalne wpływające na wahania poziomu morza

W podrozdziale tym zostanie omówiona analiza maksymalnych wezbrań sztormowych zarejestrowanych w pięciu portach znajdujących się na Morzu Bałtyckim [80]. Analiza ryzyka wystąpienia maksymalnych wezbrań sztormowych w kilku lokalizacjach jest istotnym zadaniem [81], gdyż wysokie wezbrania sztormowe mogą spowodować wystąpienie powodzi sztormowej stwarzając niebezpieczeństwo dla terenów nadmorskich. Analogicznie tworzy się modele oceny ryzyka dużych spadków wartości kilku spółek giełdowych należących do portfela inwestycyjnego. W pracy będzie poszukiwana analogia pomiędzy tymi dwoma modelami.

Przejdźmy do analizy wezbrań sztormowych zarejestrowanych w pięciu portach na Morzu Bałtyckim. Do badań użyto 72 dane z każdego portu. Dane te są maksymalnymi wezbraniem sztormowymi zarejestrowanymi w okresie 1976–2000. Lokalizację portów przedstawiono na rysunku 29 oraz wymieniono poniżej w kolejności od strony prawej na rysunku 29:

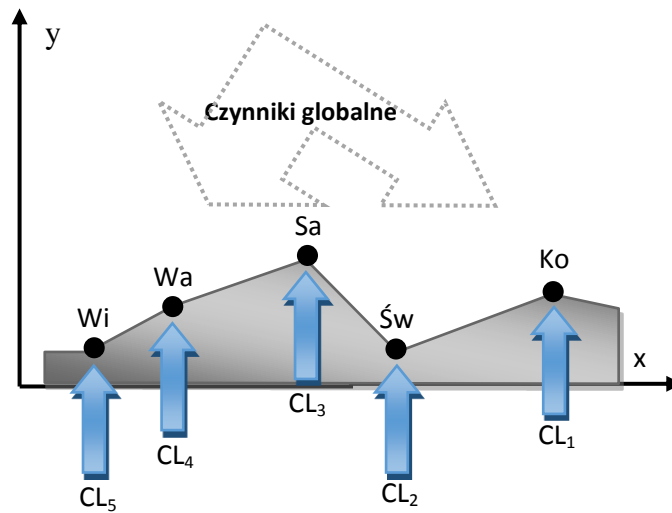
- Kołobrzeg (Ko) - port znajdujący się na otwartym morzu,
- Świnoujście (Sw) - port znajdujący się na otwartym morzu oddalony od Kołobrzegu o około 100 km,
- Sassnitz (Sa) - port na wybrzeżu niemieckim oddalony od Kołobrzegu o około 150 km (położenie przejściowe w porównaniu do portów polskich oraz pozostałych niemieckich),
- Warnemunde (Wa) - port oddalony od Kołobrzegu o około 250 km, zlokalizowany w Zatoce Meklemburskiej, częściowo osłonięty przez ląd,
- Wismar (Wi) - port oddalony od Kołobrzegu o około 300 km, również zlokalizowany w Zatoce Meklemburskiej oraz częściowo osłonięty przez ląd.



Rysunek 29. Położenie badanych portów na Morzu Bałtyckim - źródło [82].

W pracy [81] omówiono szczegóły selekcji danych oraz wybór maksymalnych wezbrań sztormowych  $L_{MAX}$  spośród danych podstawowych, to znaczy wskazań wodowskazów w poszczególnych portach. Procedura ta polegała na analizie danych podczas występowania sztormu na Morzu Bałtyckim oraz rejestracji najwyższych wskazań wodowskazów [81].

Wezbrania sztormowe powstają na skutek oddziaływania wiatrów i frontów atmosferycznych na morze. Układ ten jest w przeważającej mierze stochastyczny i jego deterministyczna analiza nie jest możliwa. Dlatego skupimy się tutaj na analizie statystycznej, podobnej do badania danych finansowych. Na rysunku 30 przedstawiono układ fizyczny podlegający badaniu [80]. Globalnym czynnikiem wpływającym na badane wielkości wezbrań sztormowych są północno - zachodnie wiatry oraz trwające przez kilkanaście godzin atmosferyczne fronty niskiego ciśnienia [81]. Lokalnie wezbrania sztormowe są również kształtowane przez lokalne czynniki [80, 81] ( $CL_1$ ,  $CL_2$ ,  $CL_3$ ,  $CL_4$ ,  $CL_5$ ), związane z różnymi lokalnymi kierunkami ruchu frontów niskiego ciśnienia oraz różniącym się lokalnie kształtem linii brzegowej. Dlatego założono, że lokalne czynniki będą się tym bardziej różnić od czynnika  $CL_1$  im bardziej będziemy się oddalać od portu w Kołobrzegu. Z tego powodu korelacja pomiędzy badaną wielkością będzie maleć wraz ze wzrostem wyżej wymienionej odległości.



Rysunek 30. Schemat układu fizycznego podlegającego badaniu.

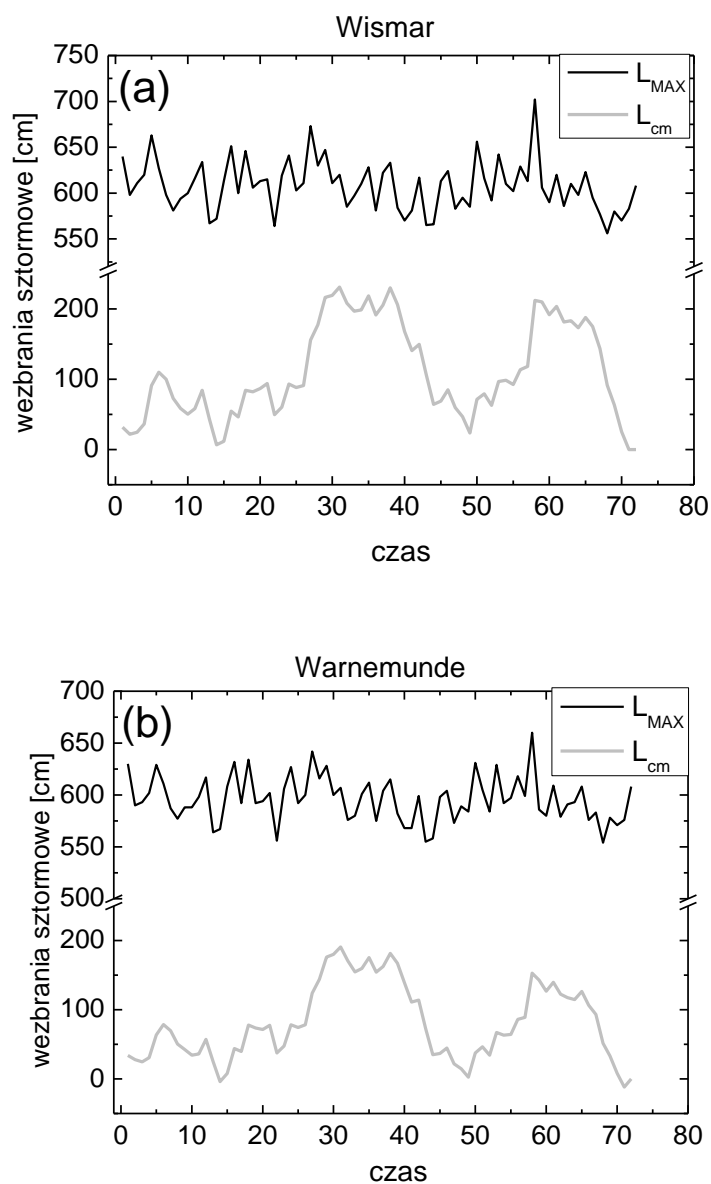
W dalszej części pracy zostanie zaprezentowana analiza auto-korelacji pojedynczych szeregów czasowych stanowiących maksymalne wezbrania sztormowe w pojedynczym porcie oraz analiza korelacji pomiędzy dwoma szeregami czasowymi stanowiącymi maksymalne wezbrania sztormowe w dwóch różnych portach.

### 5.2.2 Badania auto-korelacji

Analiza auto-korelacji zostanie przeprowadzona z wykorzystaniem formalizmu anomalnej dyfuzji oraz analizy beztrendowej, Detrended Fluctuation Analysis (DFA), dla każdego z badanych portów osobno [80]. Każdemu portowi ( $i$ ) przyporządkowano zmienną  $L_{i,MAX}$  odpowiadającą maksymalnym wezbraniom sztormowym zarejestrowanym w tym porcie. Hipotezą badawczą było sprawdzenie, czy dla danego portu występują auto-korelacje pomiędzy kolejnymi wartościami zmiennej  $L_{i,MAX}(t)$  i czy dane są w pełni stochastyczne. Do przeprowadzenia analizy zastosowano wykładnik Hursta  $H$  wyliczony z użyciem globalnej DFA [80]. Aby można było przeprowadzić tę analizę wyznaczono, wzorując się na wzorach (32) oraz (33), skumulowany sygnał w następującej postaci:

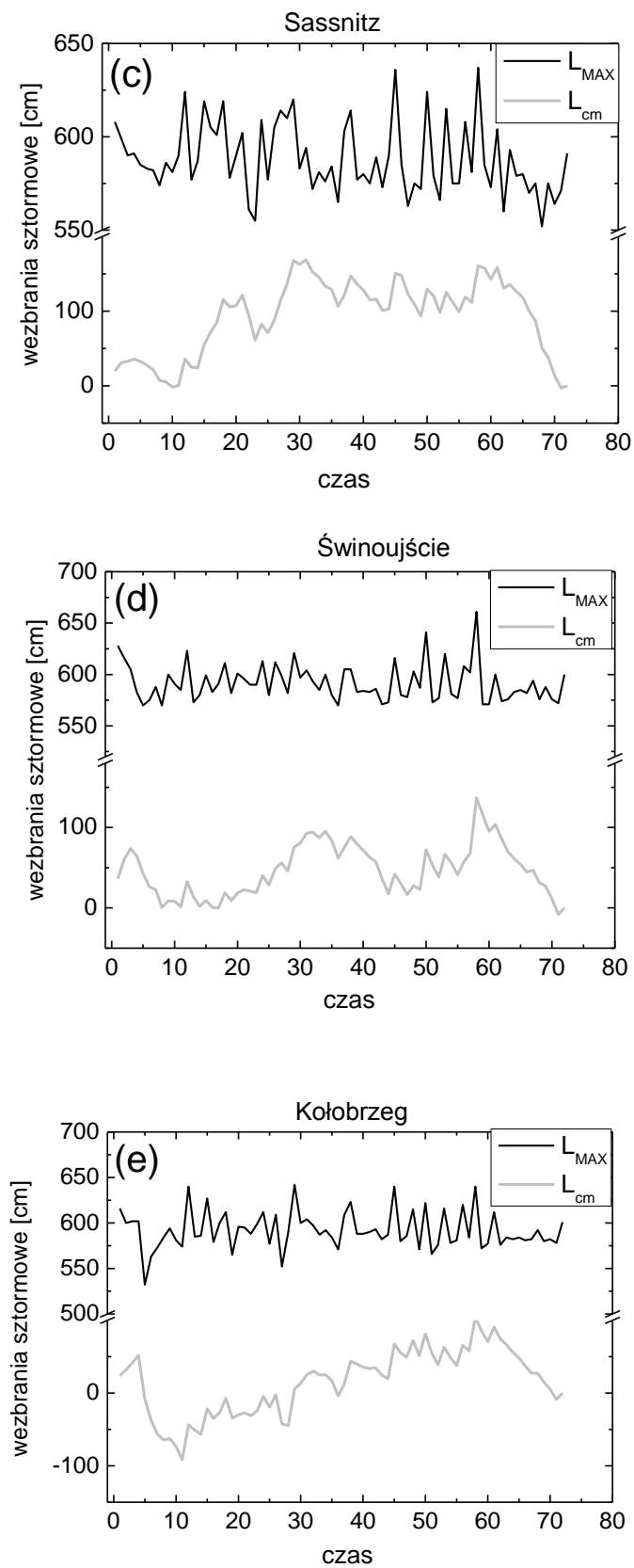
$$L_{cm}(k) = \sum_{t'=1}^k \Delta L(t') = \sum_{t'=1}^k (L(t') - \bar{L}), \quad (109)$$

gdzie  $\bar{L}$  to globalna średnia arytmetyczna sygnału, natomiast wielkość  $\Delta L(t')$  to odchylenie sygnału od globalnej średniej. Na rysunku 31 przedstawiono oryginalne sygnały  $L_{MAX}$  oraz sygnały skumulowane  $L_{cm}$ .



Rysunek 31. Szereg maksymalnych wezbrań sztormowych  $L_{MAX}$  oraz szereg skumulowany  $L_{cm}$ : (a) - Wismar, (b) - Warnemünde. Jednostką czasu są kolejne wezbrania sztormowe.





Rysunek 31 kont. Szereg maksymalnych wezbrań sztormowych  $L_{MAX}$  oraz szereg skumulowany  $L_{cm}$ :

(c) - Sassnitz, (d) - Świnoujście, (e) - Kołobrzeg.

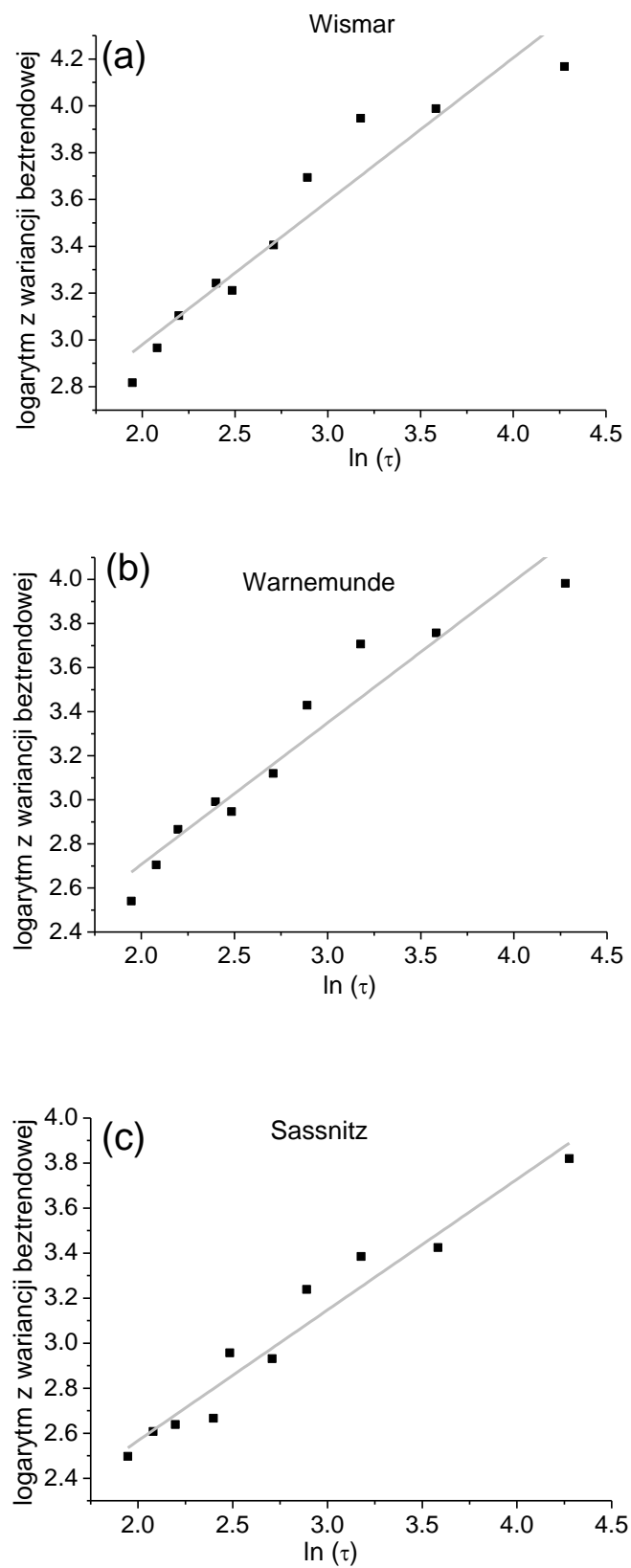
Podsumowując:

- jeżeli  $\Delta L(t') > 0$  w punkcie  $t'$  notujemy wezbranie sztormowe większe niż  $\bar{L}$  oraz wzrost wielkości  $L_{cm}$ ,
- w przeciwnym razie, jeżeli  $\Delta L(t') < 0$  w punkcie  $t'$  notujemy wezbranie sztormowe mniejsze niż  $\bar{L}$  oraz spadek wielkości  $L_{cm}$ .

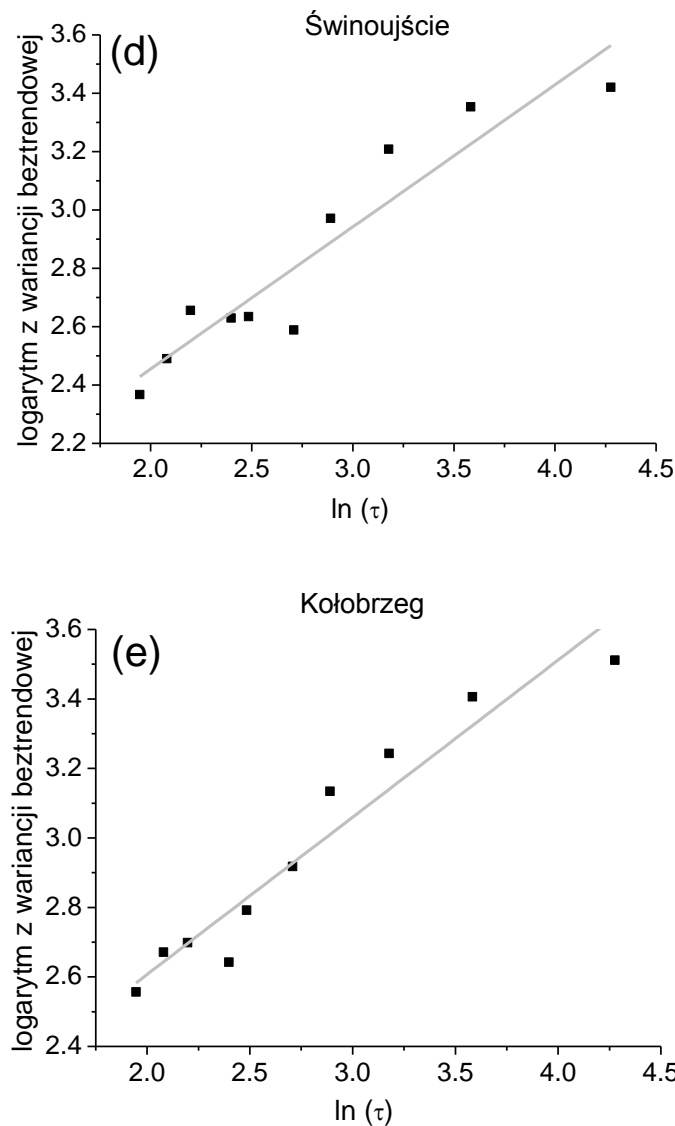
Auto-korelacja danych ekstremalnych (maksymalnych wezbrań sztormowych) prowadzi do klasteryzacji dużych lub małych wartości [64] według następujących zasad:

- dodatnia auto-korelacja - klasteryzacja dużych wartości  $\Delta L(t')$  (dużego wzrostu  $L_{cm}(k)$ ) lub małych wartości  $\Delta L(t')$  (dużego spadku  $L_{cm}(k)$ ),
- ujemna auto-korelacja - sytuacja odwrotna, w której duże i małe wartości  $\Delta L(t')$  występują naprzemiennie częściej niż gdyby były wybierane losowo.

Analiza szeregów czasowych  $L_{cm}$  może być przeprowadzona w analogii do analizy szeregu czasowego danych finansowych, gdzie dodatnia auto-korelacja odpowiada grupowaniu dużych wzrostów lub spadków, a ujemna auto-korelacja odpowiada ich naprzemiennym występowaniu. Do badania danych zastosowano standardową procedurę globalnej DFA a następnie wyznaczono wariancję beztrendową według wzorów (24) i (25). Dla każdego szeregu czasowego wyznaczono wykładnik Hursta, jako współczynnik nachylenia prostej według wzoru (27), co zaprezentowano na rysunku 32.



Rysunek 32. Dopasowanie linii prostej do wykresu logarytmu wariancji beztrendowej w funkcji logarytmu długości okna -  $\tau$ : (a) - Wismar, (b) - Warnemunde, (c) - Sassnitz.



Rysunek 32 kont. Dopasowanie linii prostej do wykresu logarytmu wariancji beztrendowej w funkcji logarytmu długości okna -  $\tau$ : (d) - Świnoujście, (e) - Kołobrzeg.

Dla omówionych danych przeprowadzono analizę „globalną”, w tym sensie, że pierwotna długość okna była równa ilości danych, czyli  $N = 72$  [80]. Dla kolejnych długości mniejszych okien, wielkości  $\tau$  były zdefiniowane następująco:  $\tau = \left\lfloor \frac{N}{k} \right\rfloor$ , gdzie  $k = 1, 2, \dots, 9, 11$  z wyłączeniem wartości  $k = 10$ , ponieważ  $\left\lfloor \frac{72}{9} \right\rfloor = \left\lfloor \frac{72}{10} \right\rfloor$ . Tabela 22 przedstawia wartości wykładnika Hursta wyliczone dla poszczególnych portów. Niepewność wartości  $H$  została wyznaczona jako niepewność dopasowania regresji liniowej.

Tabela 22. Wykładnik Hursta wyliczony dla badanych portów oraz rodzaj auto-korelacji [80].

Port	Wykładnik $H$	Niepewność wykładnika $H$	Rodzaj auto-korelacji
Wi	0.61	0.06	dodatnia
Wa	0.64	0.06	dodatnia
Sa	0.58	0.05	dodatnia
Sw	0.49	0.06	brak
Ko	0.45	0.04	ujemna

Analizując tabelę 22 można wywnioskować, że auto-korelacja rośnie, jeśli przemieszczamy się coraz bardziej na zachód. Fakt ten ma związek z coraz bardziej różniącymi się czynnikami lokalnymi, w tym prawdopodobnie z wpływem kształtu linii brzegowej, która staje się coraz bardziej urozmaicona. Aby przyjrzeć się bliżej tym danym odniesiono się do koncepcji klasteryzacji [64]. Do badania najwyższych wezbrań sztormowych - tabela 23 - zdefiniowano następujące wielkości:

- $L(q_3)$  - wielkość wezbrań sztormowych wyższych lub równych trzeciemu kwartyłowi,
- $N(q_3)$  - ilość wezbrań sztormowych wyższych lub równych trzeciemu kwartyłowi,
- $Nc(q_3)$  - ilość klasterów - przypadków następujących po sobie wezbrań sztormowych wyższych lub równych trzeciemu kwartyłowi,
- $P(Nc(q_3))$  - prawdopodobieństwo wystąpienia stwierdzonej ilości klasterów gdyby wezbrania sztormowe były szeregowane losowo,
- $E, M(N(q_3))$  - wartość przewidywana oraz modalna ilości klasterów gdyby wezbrania sztormowe były szeregowane losowo.

Tabela 23. Klasteryzacja wysokich wezbrań sztormowych.

Port	$L(q_3)$ [cm]	$N(q_3)$	$Nc(q_3)$	$P(Nc(q_3))$	$E(N(q_3))$	$M(N(q_3))$
Wi	622.55	18	3	19.6%	4.3	4
Wa	608	20	4	18.5%	5.3	5
Sa	602.55	18	5	21.7%	4.3	4
Sw	601.55	18	5	21.7%	4.3	4
Ko	602	19	2	6.1%	4.7	5

Analizując tabelę 23 można stwierdzić, iż ilość klasterów wysokich wezbrań sztormowych dla Kołobrzegu jest niewielka - równa 2, przy wartości przewidywanej równej 4.7. Świadczy to o anty-korelacji wysokich wezbrań sztormowych dla Kołobrzegu - po zarejestrowanym wysokim wezbraniu sztormowym kolejne wysokie wezbranie jest mniej prawdopodobne.

Analogicznie do badania niższych wezbrań sztormowych - tabela 24 - zdefiniowano następujące wielkości:

- $L(q_1)$  - wielkość wezbrań sztormowych niższych lub równych pierwszemu kwartylowi.
- $N(q_1)$  - ilość wezbrań sztormowych niższych lub równych pierwszemu kwartylowi.
- $Nc(q_1)$  - ilość klasterów - przypadków następujących po sobie wezbrań sztormowych niższych lub równych pierwszemu kwartylowi.
- $P(Nc(q_1))$  - prawdopodobieństwo wystąpienia stwierdzonej ilości klasterów gdyby wezbrania sztormowe były szeregowane losowo.
- $E, M(N(q_1))$  - wartość przewidywana oraz modalna ilości klasterów gdyby wezbrania sztormowe były szeregowane losowo.

Tabela 24. Klasteryzacja niskich wezbrań sztormowych.

Port	$L(q_1)$ [cm]	$N(q_1)$	$Nc(q_1)$	$P(Nc(q_1))$	$E(N(q_1))$	$M(N(q_1))$
Wi	585.45	18	8	1.7%	4.3	4
Wa	580	19	7	9.3%	4.7	5
Sa	575	21	9	4.5%	5.8	6
Sw	578.9	18	6	13.1%	4.3	4
Ko	580	19	4	22.5%	4.7	5

Analizując tabelę 24 można stwierdzić, iż ilość klasterów niskich wezbrań sztormowych dla portów Wismar, Wardemunde i Sassnitz jest duża, równa odpowiednio 8, 7 i 9, przy wartościach przewidywanych równych odpowiednio 4.3, 4.7 oraz 5.8. Świadczy to o auto-korelacji niskich wezbrań sztormowych dla tych trzech portów. Zjawisko to jest szczególnie silne dla Wismar, gdyż prawdopodobieństwo jego wystąpienia jest równe jedynie 1.7%. Dla trzech niemieckich portów można stwierdzić, iż po zarejestrowaniu niskiego wezbrania sztormowego kolejne niskie wezbranie sztormowe jest bardziej prawdopodobne. W tym miejscu nasuwają się następujące wnioski:

- badając wykładnik Hursta dla danych z Kołobrzegu stwierdzono ujemną auto-korelację, która odzwierciedla się w fakcie, iż po zarejestrowaniu wysokiego wezbrania sztormowego kolejne wysokie wezbranie sztormowe jest mniej prawdopodobne,
- badając wykładnik Hursta dla danych ze Świnoujścia stwierdzono brak auto-korelacji,
- badając wykładnik Hursta dla Wismar, Wardemunde i Sassnitz stwierdzono dodatnie auto-korelacje, które odzwierciedlają się w fakcie, iż po zarejestrowaniu niskiego wezbrania sztormowego kolejne niskie wezbranie sztormowe jest bardziej prawdopodobne.

Analizując auto-korelacje maksymalnych wezbrań sztormowych otrzymano pewne wskazówki związane z przewidywaniem wezbrań sztormowych oraz oceną ryzyka wystąpienia bardzo wysokich wezbrań sztormowych [80]. Istotnie zjawisko zmiennej auto-korelacji występuje również dla danych finansowych, często jednak jest ona analizowana nie w obrębie kilku zmiennych, ale jednej zmiennej, czyli kursów pojedynczego papieru wartościowego [2]. Aby ocenić faktyczne ryzyko wystąpienia wysokich wezbrań sztormowych w kilku lokalizacjach należy przeprowadzić analizę korelacji z użyciem funkcji Kopuły, co zostanie zrobione w następnym rozdziale. Podobna procedura zostanie również zaproponowana dla danych finansowych.

### 5.2.3 Analiza danych z wykorzystaniem funkcji Kopuły

Do analizy korelacji wykorzystano funkcje Kopuły [80]. W pierwszej kolejności do danych ekstremalnych dopasowano rozkłady brzegowe. Do danych ze Świnoujścia nieposiadających auto-korelacji (tabela 22) dopasowano rozkład RWE opisany w rozdziale 4.1. Do pozostałych danych posiadających dodatnie lub ujemne auto-korelacje obok rozkładu RWE zastosowano rozkład  $RJS_U$  oraz RL, co zostało opisane w rozdziale 4.2 [80]. W trakcie analizy używano testu Kołmogorowa - Smirnowa do oceny jakości dopasowania [80]. W teście, wartość  $d$  to supremum różnicy pomiędzy danymi oraz badanym rozkładem prawdopodobieństwa, natomiast wartości  $p$  to prawdopodobieństwo, że dane są opisane przez badany rozkład. W tabeli 25, 26 oraz 27 przedstawiono parametry dopasowania rozkładów RWE,  $RJS_U$  oraz RL.

Tabela 25. Parametry oraz jakość dopasowania rozkładu RWE.

	Parametry rozkładu RWE			Parametry testu S-K	
Port	$\xi$	$\mu$	$a$	$d$	$p$
(Ko)	-0.207	584.3	19.6	0.103	0.400
(Św)	0	583.9	12.8	0.089	0.605
(Sa)	0	579.2	15.7	0.067	0.905
(Wa)	-0.173	587.4	20.3	0.051	0.992
(Wi)	-0.107	596.5	24.3	0.083	0.709

Tabela 26. Parametry oraz jakość dopasowania rozkładu  $RJS_U$ .

	Parametry rozkładu $RJS_U$				Parametry testu S-K	
Port	$\zeta$	$\nu$	$\mu$	$a$	$d$	$p$
(Ko)	-0.683	1.22	579.2	16.3	0.063	0.938
(Sa)	-8.66	3.65	520.6	12.2	0.068	0.89
(Wa)	-12.67	7.07	450.8	49.2	0.053	0.986
(Wi)	-10.6	4.70	483.0	26.0	0.083	0.709

Tabela 27. Parametry oraz jakość dopasowania rozkładu LD.

	Parametry rozkładu RL		Parametry testu S-K	
Port	$\mu$	$a$	$d$	$p$
(Ko)	90.7	11.2	0.101	0.427
(Sa)	586.5	10.8	0.090	0.585
(Wa)	596.3	12.4	0.067	0.907
(Wi)	607.0	15.3	0.062	0.942

Analizując tabelę 25 można stwierdzić, że rozkład RWE najgorzej modeluje dane (Ko), posiadające ujemną auto-korelację, natomiast stosunkowo dobrze modeluje pozostałe dane, które nie posiadają auto-korelacji - (Św) - oraz te, które posiadają dodatnią auto-korelację: (Wi), (Wa), (Sa) [80]. Badając tabelę 26 można stwierdzić, że rozkład  $RJS_U$  dobrze modeluje dane posiadające zarówno dodatnią jak i ujemną auto-korelację [80]. Dane umieszczone w tabeli 27 pokazują, że rozkład RL najlepiej

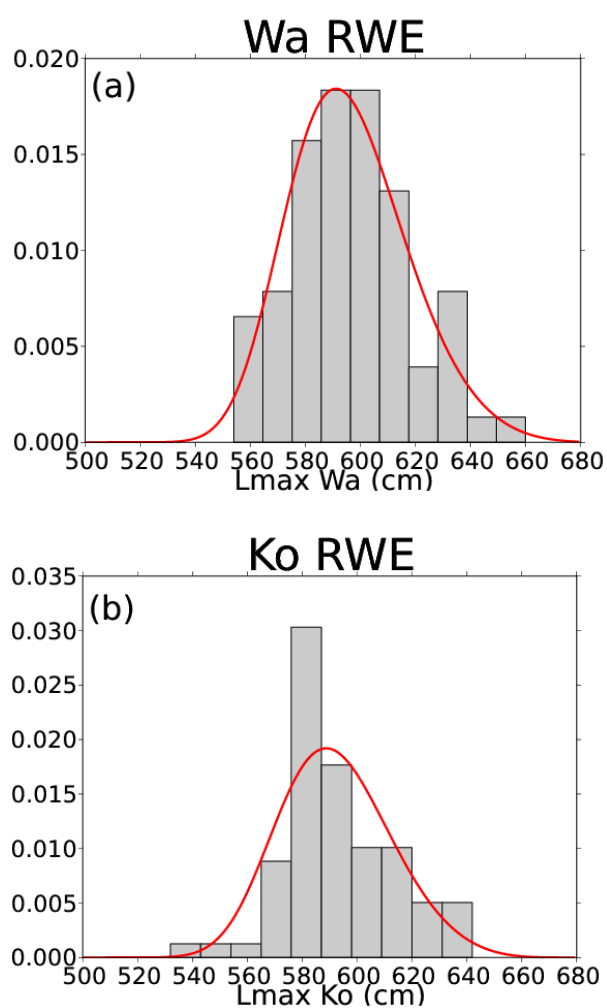


modeluje dane (Wa) oraz (Wi) posiadające wyraźną dodatnią auto-korelację [80]. W tabeli 28 przedstawiono rozkłady prawdopodobieństwa w kolejności jakości dopasowania.

Tabela 28. Rozkłady prawdopodobieństwa w kolejności jakości dopasowania - od strony lewej.

port	rozkład	$p$	rozkład	$p$	rozkład	$p$
(Ko)	RJS <sub>U</sub>	0.938	RL	0.427	RWE	0.400
(Św)	RWE	0.605				
(Sa)	RWE	0.905	RJS <sub>U</sub>	0.89	RL	0.585
(Wa)	RWE	0.992	RJS <sub>U</sub>	0.986	RL	0.907
(Wi)	RL	0.942	RWE	0.709	RJS <sub>U</sub>	0.709

Na rysunku 33 przedstawiono jako przykład najgorsze i najlepsze dopasowanie rozkładu RWE.



Rysunek 33. Rozkład RWE - (a) najlepsze oraz (b) najgorsze dopasowanie.

Aby przeprowadzić analizę korelacji, użyto portu w Kołobrzegu jako portu referencyjnego [80]. Dane zarejestrowane w kolejnych portach porównywano z danymi zarejestrowanymi w Kołobrzegu, dzięki czemu otrzymano 4 pary danych, które zostały przedstawione poniżej w kolejności odpowiadającej malejącej korelacji. Założono, że maleje ona wraz z wzrostem odległości pomiędzy portami:

- $[L_{MAX, Ko}, L_{MAX, Sw}]$  - Kołobrzeg (Ko) i Świnoujście (Sw) ,
- $[L_{MAX, Ko}, L_{MAX, Sa}]$  - Kołobrzeg (Ko) i Sassnitz (Sa) ,
- $[L_{MAX, Ko}, L_{MAX, Wa}]$  - Kołobrzeg (Ko) i Warnemunde (Wa) ,
- $[L_{MAX, Ko}, L_{MAX, Wi}]$  - Kołobrzeg (Ko) i Wismar (Wi).

Aby potwierdzić malejącą korelację wraz ze wzrostem odległości od Kołobrzegu w tabeli 29 zestawiono współczynniki korelacji rang  $\tau$  Kendall'a. Malejąca korelacja wynika z malejącego podobieństwa czynników lokalnych, gdyż wraz z odległością dynamika lokalnych frontów niskiego ciśnienia coraz bardziej się różni. Analizując tabelę 29 można stwierdzić niezerową dodatnią korelację pomiędzy najbardziej oddalonymi portami, świadczącą o występowaniu czynnika globalnego wpływającego na wszystkie porty. Czynnikiem tym są północno - zachodnie wiatry oraz trwające przez kilkanaście godzin atmosferyczne fronty niskiego ciśnienia [81].

Tabela 29. Pary portów oraz współczynnik korelacji rang.

Para portów	$\tau$ Kendall'a
(Sw) oraz (Ko)	0.596
(Sa) oraz (Ko)	0.445
(Wa) oraz (Ko)	0.261
(Wi) oraz (Ko)	0.240

Zakładając zbliżoną dynamikę poszczególnych zmiennych  $L_{MAX}$ , można wprowadzić modele posiadające ten sam rozkład brzegowy dla każdego składnika danej pary, proponując:

- rozkład RWE jako naturalny wybór na rozkład danych ekstremalnych - dla wszystkich czterech par,
- rozkład  $RJS_U$  dla pary  $[L_{MAX, Ko}, L_{MAX, Sa}]$  oraz  $[L_{MAX, Ko}, L_{MAX, Wa}]$ , gdyż według tabeli 28 rozkład ten najlepiej modeluje dane dla Kołobrzegu oraz bardzo dobrze dane dla Sassnitz i Warndemunde,
- rozkład RL dla pary  $[L_{MAX, Ko}, L_{MAX, Wa}]$ , gdyż według tabeli 28 rozkład ten najlepiej modeluje dane dla portu Wismar.

Do danych tych dopasowano funkcje Kopuły omówione w rozdziale 3.2, według procedury z użyciem współczynnika korelacji  $\tau$  Kendall'a, która została omówiona w rozdziale 3.4.4. Przypomnijmy, w przypadku Kopuły Archimedesowskich parametr  $\theta$ , tak samo jak współczynnik korelacji rang  $\tau$  Kendall'a, nie zależy od rozkładów brzegowych. W przypadku Kopuły Studenta natomiast parametr  $\rho$ , jako wyznaczony na podstawie współczynnika  $\tau$  Kendall'a, nie zależy do parametrów rozkładów brzegowych, ale parametr  $\nu$  wyznaczony z użyciem funkcji maksymalnej wiarygodności wykazuje tę zależność. Następnie, wyznaczono wartości indeksów górnej zależności ogonowej dla funkcji Kopuły (tabela 30) oraz danych empirycznych (tabela 31), przy założeniu  $\alpha = 0.9$ , z użyciem wzoru (73). Użyto przy tym stosunkowo niskiej wartości  $\alpha = 0.9$  ze względu na małą ilość danych (72) poddanych analizie [80].

Tabela 30. Indeksy górnej zależności ogonowej dla Kopuły - w przypadku Kopuły Studenta różne wartości odpowiadają różnym parametrom  $\nu$  dla różnych rozkładów brzegowych.

Para portów	Kopuła		
	odw. Claytona	Gumbela	Studenta
(Św) oraz (Ko)	0.79	0.68	0.24
(Sa) oraz (Ko)	0.65	0.53	0.40
(Wa) oraz (Ko)	0.38	0.33	0.16-0.20
(Wi) oraz (Ko)	0.33	0.31	0.19-0.25

Tabela 31. Empiryczne wartości indeksu górnej zależności ogonowej

Para portów	RWE	RL	RJS <sub>U</sub>
(Św) oraz (Ko)	0.65		
(Sa) oraz (Ko)	0.88		0.88
(Wa) oraz (Ko)	0.38		0.38
(Wi) oraz (Ko)	0.38	0.40	

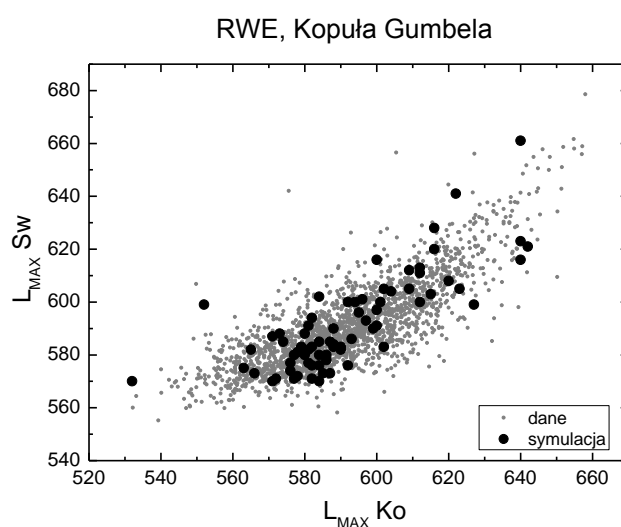
Badanie indeksu górnej zależności ogonowej dla maksymalnych wezbrań sztormowych odzwierciedla prawdopodobieństwo wystąpienia wartości ekstremalnych w kilku miejscach, co można powiązać z niebezpieczeństwem wystąpienia powodzi sztormowej na dużą skalę. Analizując tabelę 31 można stwierdzić, iż indeks górnej zależności ogonowej występuje również dla odległych portów sugerując

możliwość jednoczesnego wystąpienia bardzo wysokich wezbrań sztormowych. Zjawisko to jest analogiczne do dynamiki zachowań danych finansowych, gdzie istnieje ryzyko jednoczesnego załamania się kursów istotnie różniących się papierów wartościowych. Analizując tabele 30 oraz 31 można stwierdzić, że najlepszy model indeksu górnej zależności ogonowej dostarcza Kopuła Gumbela dla pary  $[L_{MAX, Ko}, L_{MAX, Sw}]$  oraz odwrócona Kopuła Clayтона dla pozostałych par. W tabeli 32 przedstawiono parametry najbardziej pasujących funkcji Kopuła [80].

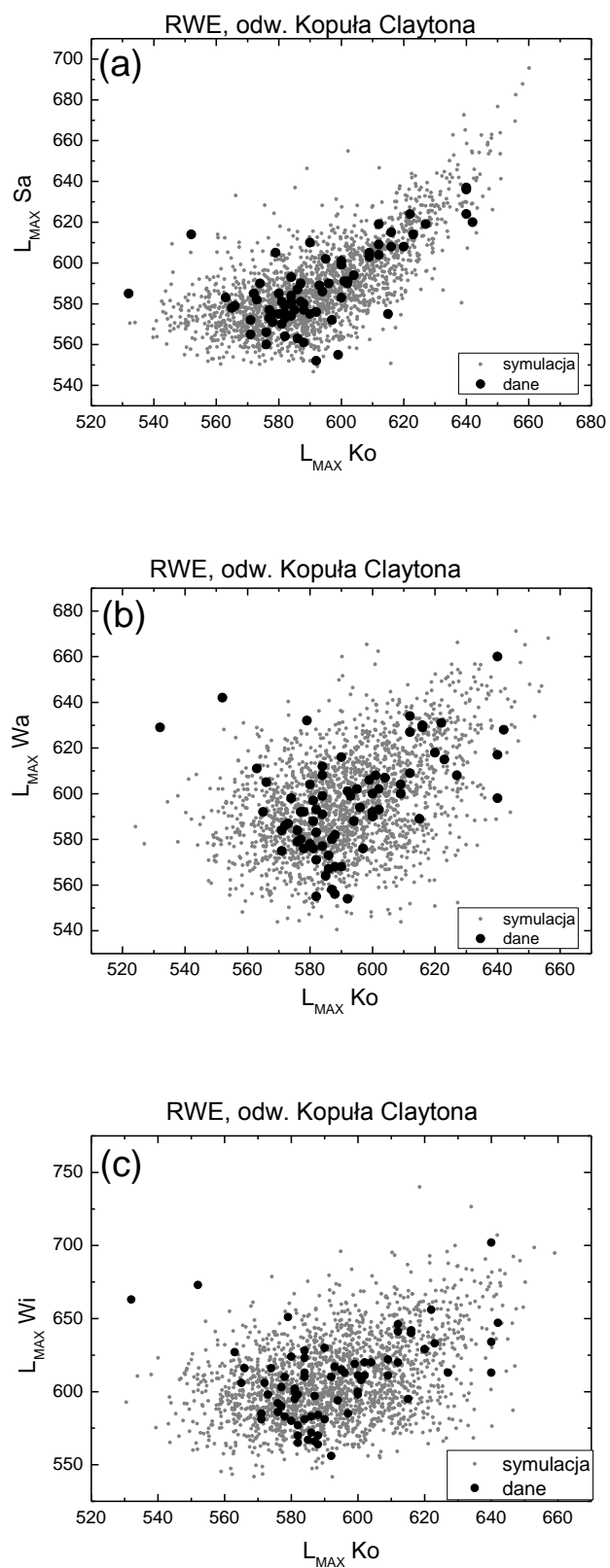
Tabela 32. Parametr  $\theta$  dopasowany do funkcji Kopuła.

Para portów	Kopuła	
	odw. Clayтона	Gumbela
(Sw) oraz (Ko)	2.95	2.48
(Sa) oraz (Ko)	1.60	1.80
(Wa) oraz (Ko)	0.71	1.35
(Wi) oraz (Ko)	0.63	1.32

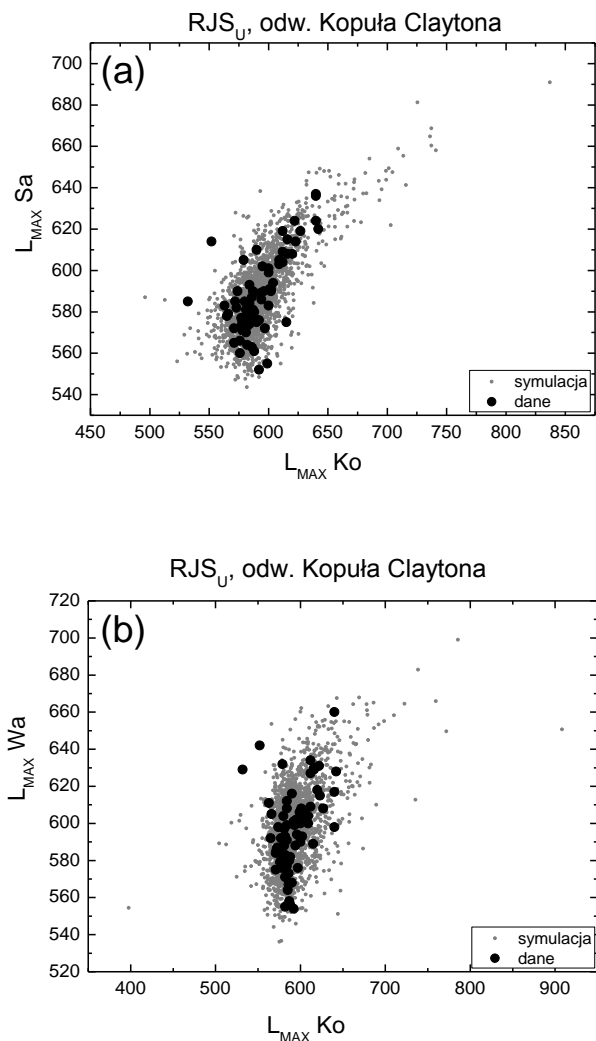
Na rysunkach 34 - 37 przedstawiono dane empiryczne - zarejestrowane maksymalne wezbrania sztormowe oraz dane wygenerowane w przestrzeni wezbrań sztormowych przez wybrane podczas badań funkcje Kopuły oraz rozkłady brzegowe [80].



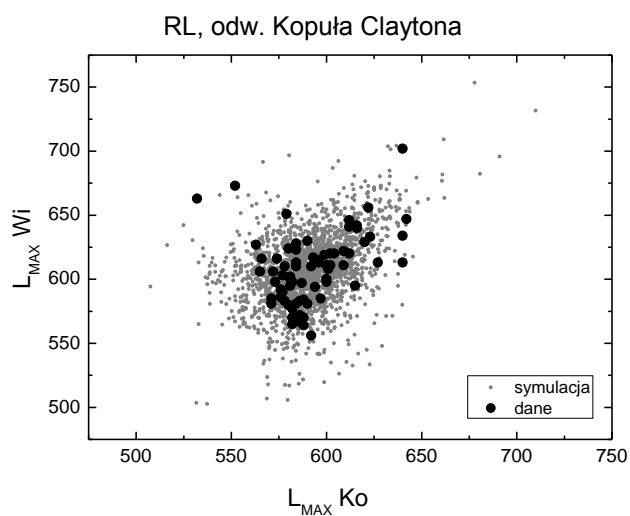
Rysunek 34. Dane empiryczne oraz dane zasymulowane w przestrzeni wezbrań sztormowych, rozkład RWE, Kopuła Gumbela - dane (Ko) oraz ( $\acute{S}$ Sw).



Rysunek 35. Dane empiryczne oraz dane zasymulowane w przestrzeni wezbrań sztormowych, rozkład RWE, odwrócona Kopuła Claytona, dane (a) - (Ko) oraz (Sa), (b) - (Ko) oraz (Wa) i (c) - (Ko) oraz (Wi).



Rysunek 36. Dane empiryczne oraz dane zasymulowane w przestrzeni wezbrań sztormowych, rozkład  $RJS_U$ , odwrócona Kopuła Claytona, dane: (a) - (Ko) oraz (Sa) i (b) - (Ko) oraz (Wa).



Rysunek 37. Dane empiryczne oraz dane zasymulowane w przestrzeni wezbrań sztormowych, rozkład RL, odwrócona Kopuła Claytona - dane (Ko) oraz (Wi).

Dokonując analizy wizualnej rysunków 34 - 37 można zaproponować następujące modele:

- dane (Ko) oraz (Sw) - rozkład RWE, Kopuła Gumbela,
- dane (Ko) oraz (Sa) i (Ko) oraz (Wa) - rozkład RWE, odwrócona Kopuła Claytona,
- dane (Ko) oraz (Wi) - rozkład RL, odwrócona Kopuła Claytona.

Indeksy górnej zależności ogonowej wyznaczone za pomocą parametrów modelu pozwalają na oszacowanie ryzyka wystąpienia powodzi sztormowej w kilku portach jednocześnie. Jest to istotne w przypadku modelowania zabezpieczeń przeciwko powodziom sztormowym. Analizę tą można rozwinąć analizując dodatkowo auto-korelacje wezbrań sztormowych. W rozdziale 5.2.2 stwierdzono klasteryzację niskich wezbrań sztormowych dla danych z portów (Wi), (Sa) oraz (Wa), jak również przeciwne zjawisko do klasteryzacji, dla wysokich wezbrań sztormowych dla danych z portu (Ko). Dlatego w przypadku zarejestrowania niskiego wezbrania sztormowego w jednym z niemieckich portów, kolejne niskie wezbranie sztormowe jest bardziej prawdopodobne, co może wskazywać (lokalnie w czasie) na model Kopuły o niższej górnej zależności ogonowej (np. Kopułę Gumbela). Podobnie, jeśli zarejestrujemy wysokie wezbranie sztormowe w Kołobrzegu, kolejne wysokie wezbranie sztormowe jest mniej prawdopodobne, co można uwzględnić również wybierając Kopułę o niższej górnej zależności ogonowej [80]. W bieżącym rozdziale omówiono model oceny ryzyka wykorzystujący funkcje Kopuły oraz zastosowano dodatkowo analizę auto-korelacji, która wspiera procedurę wyboru właściwej funkcji Kopuły określającej ryzyko. Podobnie funkcje Kopuły wykorzystuje się do analizy ryzyka jednoczesnego spadku wartości wielu papierów wartościowych [1]. W następnych rozdziałach zostanie omówiony analogiczny model oceny ryzyka dla papierów wartościowych, jedynie z tą różnicą, że badane były maksymalne ujemne stopy zwrotu, a więc dane minimalne. Analiza przebiega podobnie, jedynie należy zamienić indeks górnej zależności ogonowej na indeks dolnej zależności ogonowej, Kopułę Gumbela na odwróconą Kopułę Gumbela oraz odwróconą Kopułę Claytona na Kopułę Claytona.

### 5.3. Badanie danych finansowych

#### 5.3.1 Badanie danych ekstremalnych

W kolejnym rozdziale, na bazie doświadczeń wynikających z analizy zjawisk geofizycznych, zostaną omówione dane finansowe, które były badane w celu ewaluacji modelu oceny ryzyka ekstremalnych spadków ich wartości. W rozdziale 3.3 pokazano, iż wykładnik dyfuzji (Hursta  $H$ ) obliczony z użyciem lokalnego DFA może być użyty jako wskaźnik przewidujący załamanie się kursów

akcji lub zmianę trendu. Dokładnie ujmując, sygnałem zbliżającej się zmiany trendu lub załamania jest spadek wykładnika Hursta poniżej progowej wielkości  $H_{threshold}$  takiej, że  $H_{threshold} < 0.5$ . W pracy [83] wykorzystano kombinację analizy anomalnej dyfuzji oraz funkcji Kopuły do oceny ryzyka inwestycji w akcje spółek notowanych na GPW w Warszawie. W opisaney tam analizie badano kursy akcji siedmiu spółek notowanych na GPW w Warszawie, należących do indeksu WIG 20, w tym czterech banków (PEKAO, PKO BP, BZ WBK, MBANK, HANDLOWY) oraz trzech spółek niebędących bankami – PKN ORLEN, KGHM oraz ASSECO.

Bank PEKAO oraz PKO BP to duże banki notowane na GPW w Warszawie - ich udział w indeksie WIG 20 na dzień 02.01.2014 był równy odpowiednio 13.6% oraz 16.0%. Kolejne banki (BZ WBK, MBANK oraz HANDLOWY) zostały uszeregowane w kolejności malejącego udziału w indeksie WIG 20, równego odpowiednio 6.2%, 3.6% oraz 2.0%. Kolejne trzy badane spółki nie należą do sektora bankowego i zostały wybrane w taki sposób, aby reprezentować różne sektory oraz różne udziały w indeksie WIG 20. Ich wybór uzasadnia się faktem, że tworzenie portfeli inwestycyjnych ze spółek słabiej skorelowanych, należących do różnych sektorów, jest istotnym wyzwaniem dla inwestorów. Spółka KGHM reprezentuje sektor wydobywczy, a jej udział w indeksie WIG 20 równy 8.6% ocenia się jako relatywnie duży. Spółka PKN ORLEN reprezentuje sektor paliwowy, a jej udział w indeksie WIG 20 równy 6.9% ocenia się jako średni - porównywalny z bankiem BZ WBK. Natomiast spółka ASSECO reprezentuje sektor oprogramowania, a jej udział w indeksie WIG 20 równy 1.9% ocenia się jako niski - porównywalny z bankiem HANDLOWY. Warto w tym miejscu przypomnieć czynniki globalne i lokalne kształtujące dynamikę kursów akcji:

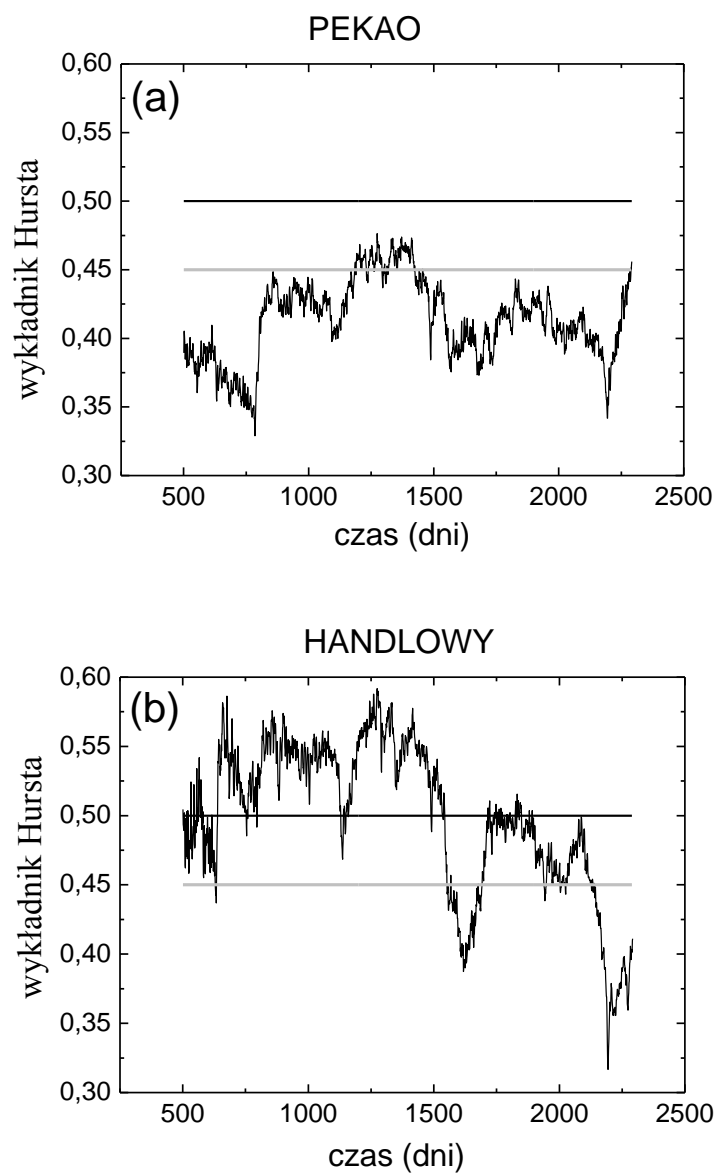
- czynniki globalne - koniunktura makroekonomiczna, ogólny stan rynku (rosnący lub spadkowy) i kolektywna strategia inwestorów, sytuacja polityczna, relacje międzypaństwowe,
- czynniki lokalne - zysowność i perspektywy poszczególnych spółek, zysowność i perspektywy poszczególnych sektorów, indywidualne strategie inwestorów.

Z faktu tego wynika, iż podobieństwo branż oraz podobieństwo wielkości spółek będzie odzwierciedlane w podobieństwie czynników lokalnych oraz silniejszej korelacji pomiędzy kursami akcji.

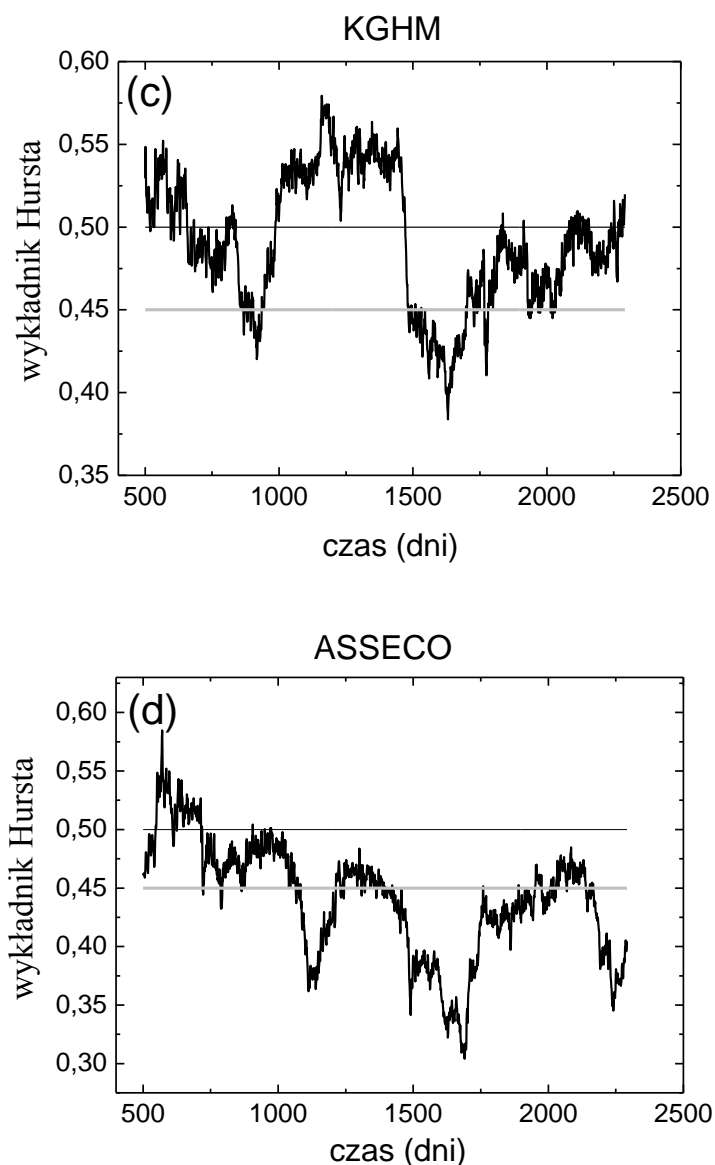
Podczas badań połączono koncepcję wykładnika Hursta wyznaczonego z użyciem lokalnego DFA, używanego do przewidywania ekstremalnych wartości stóp zwrotu (ekstremalnych spadków kursów akcji) z koncepcją funkcji Kopuły używaną do oceny ryzyka jednoczesnego wystąpienia ekstremalnych wartości stóp zwrotu [83] dla kilku papierów wartościowych. W celu wyliczenia wykładnika Hursta zastosowano procedurę opisaną przez wzory (24), (25) i (27), w tym zastosowano okno obserwacyjne o długości  $N = 500$ , w którym badano dzienne kursy akcji. Szczegóły opisano



w pracy [83]. Przykłady wartości wykładnika Hursta wyznaczone dla wybranych spółek: PEKAO, HANDLOWY, KGHM oraz ASSECO przedstawiono na rysunku 38.



Rysunek 38. Wartości wykładnia Hursta wyliczone dla wybranych spółek: (a) - PEKAO, (b) - HANDLOWY.



Rysunek 38 kont. Wartości wykładnika Hursta wyliczone dla wybranych spółek: (c) - KGHM, (d) - ASSECO.

Wartość wykładnika Hursta wyliczono za okres od 03.11.2006 do 02.01.2014 [83]. Analizując rysunek 38 można znaleźć pewną analogię pomiędzy wykładnikiem Hursta wyliczonym dla kursów akcji oraz wyliczonym dla zmian koncentracji jonów sodu i magnezu w rdzeniach arktycznych - rysunek 21. Patrząc z drugiej strony, duża różnica w dynamice wykładnika Hursta pomiędzy PEKAO i KGHM lub ASSECO może świadczyć o dużej różnicy czynników lokalnych.

Dalsza analiza obejmuje pary papierów wartościowych składające się na dwuskładnikowe portfele inwestycyjne [83]. Pierwszym składnikiem każdego portfela jest spółka PEKAO, natomiast drugim spółka będąca lub niebędąca bankiem. W pracy [83] zaproponowano wartość progową

$H_{threshold} = 0.45$ , poniżej której rejestruje się sygnał anty-korelacji. Dlatego założono, że jeśli dla obu składników portfela  $H > H_{threshold}$  to ich interakcja jest gaussowska, a inwestycja bezpieczniejsza [83]. W przypadku niespełnienia warunku założono, że występuje istotnie prawdopodobieństwo znacznego spadku cen obu papierów wartościowych należących do omówionego portfela.

W celu porównania kursów akcji można wprowadzić dla  $i$ -tej spółki procentową stopę zwrotu  $R(t)_i$  w czasie  $t$  opisaną wzorem:

$$R(t)_i = \frac{P(t+1)_i - P(t)_i}{P(t)_i}. \quad (110)$$

Badaniom poddano ujemne stopy zwrotu mniejsze od zadanej wartości progowej  $\alpha$ , przy czym założono  $\alpha = -1\%$  [83]. Następnie utworzono zmienne dwuwymiarowe  $[R_{PEKAO}(t'), R_j(t')]$ , gdzie  $R_{PEKAO}(t') = R_{PEKAO}(t)$  oraz  $R_j(t') = R_j(t)$  jeżeli  $R_{PEKAO}(t) < \alpha$  oraz  $R_j(t) < \alpha$ . Procedura została przeprowadzona każdorazowo dla każdego  $j$  numerującego kolejne spółki należące do drugiego składnika portfela inwestycyjnego [83].

### 5.3.2 Modele typu VaR

W analizie danych ekonomicznych warto odnieść się do modeli statystycznych stosowanych w finansach do analizy ryzyka inwestycyjnego. Są to modele typu „VaR” - „Value at Risk” - używane w celu oszacowania maksymalnej straty, którą można ponieść inwestując w papiery wartościowe przy założeniu ustalonego poziomu ufności [2, 34, 35]. Klasyczne podejście zakładało wyznaczenie wartości narażonej na ryzyko z wykorzystaniem wielowymiarowego rozkładu Gaussa lub empirycznej dystrybuanty wielowymiarowej. Podejście to jednak nie zawsze się sprawdzało, ponieważ rozkład Gaussa nie jest dobrym modelem dla danych finansowych, natomiast zastosowanie wielowymiarowej dystrybuanty empirycznej wymaga dużych mocy obliczeniowych [2]. Rozkład Gaussa nie opisuje dobrze danych finansowych z powodu występujących auto-korelacji oraz dynamiki zbliżonej do procesów Levy'ego [2]. Dlatego w pracy autora dysertacji [83] zaproponowano wykorzystanie funkcji Kopuły do utworzenia wielowymiarowego rozkładu prawdopodobieństwa modelującego dane finansowe oraz wyznaczającego prawdopodobieństwa (ryzyka) wystąpienia wartości ekstremalnych [1]. Należy też zwrócić od razu uwagę na fakt, że stwierdzony charakter auto-korelacji danych może być pomocny w wyborze właściwej funkcji Kopuły. Z tego powodu analizę danych finansowych przeprowadzoną z wykorzystaniem funkcji Kopuły uzupełnia się poprzez analizę auto-korelacji [83].

### 5.3.3 Modelowanie kursów akcji spółek giełdowych notowanych na GPW w Warszawie

Analizę korelacji przeprowadzono z użyciem funkcji Kopuły dla par szeregów czasowych mających ujemne stopy zwrotu, opisane wcześniej w rozdziale 5.3.1. W tym celu do każdej zmiennej w parze dopasowano jednowymiarowe rozkłady brzegowe - badano rozkład wartości ekstremalnych - RWE oraz rozkład  $RJS_U$ . W tabeli 33 i 34 opisano jakość dopasowania. Do analizy wybrano jednowymiarowe rozkłady prawdopodobieństwa posiadające wyższe wartości  $p$  wynikające z testu Kołmogorowa - Smirnowa [83]. Analiza wymagała zastosowania różnych rozkładów brzegowych wewnątrz niektórych par.

Tabela 33. Jakość dopasowania rozkładów brzegowych - pary banków.

Spółki	Rozkład	Wartość - p PEKAO	Wartość - p drugi bank
PEKAO PKO BP	RWE	<b>0.540</b>	0.070
	$RJS_U$	0.111	<b>0.211</b>
PEKAO BZ WBK	RWE	<b>0.539</b>	$8.7 \cdot 10^{-5}$
	$RJS_U$	0.137	<b>0.573</b>
PEKAO HANDLOWY	RWE	<b>0.361</b>	0.113
	$RJS_U$	0.194	<b>0.305</b>
PEKAO MBANK	RWE	<b>0.720</b>	0.144
	$RJS_U$	0.226	<b>0.709</b>

Tabela 34. Jakość dopasowania rozkładów - para bank i spółka niebędąca bankiem.

Spółki	Rozkład	Wartość - p PEKAO	Wartość - p druga spółka
PEKAO KGHM	RWE	<b>0.215</b>	0.448
	$RJS_U$	0.051	<b>0.461</b>
PEKAO PKN ORLEN	RWE	<b>0.863</b>	0.002
	$RJS_U$	0.222	<b>0.838</b>
PEKAO ASSECO	RWE	<b>0.357</b>	<b>0.925</b>
	$RJS_U$	0.152	0.246

Dokładną dyskusję wyników przeprowadzono w [83]. W dalszej analizie założono, że podobieństwo branży bardziej wpływa na podobne zachowanie się kursów akcji niż podobieństwo wielkości spółek - jest to ogólne założenie stosowane przy optymalnym wyborze portfeli inwestycyjnych. Poniżej zaprezentowano zmienne dwuwymiarowe podlegające analizie w kolejności odzwierciedlającej malejące podobieństwo danych:

## Analiza danych empirycznych

- $[R_{PEKAO}, R_{PKO BP}]$ ,
- $[R_{PEKAO}, R_{BZ WBK}]$ ,
- $[R_{PEKAO}, R_{MBANK}]$ ,
- $[R_{PEKAO}, R_{HANDLOWY}]$ ,
- $[R_{PEKAO}, R_{KGHM}]$ ,
- $[R_{PEKAO}, R_{PKN ORLEN}]$ ,
- $[R_{PEKAO}, R_{ASSECO}]$ .

W tabeli 35 pokazano wartości współczynnika korelacji rang  $\tau$  Kendall'a, dla poszczególnych par danych.

Tabela 35. Współczynnik korelacji rang dla poszczególnych danych.

Para danych	$\tau$ Kendall'a	Udział w WIG 20	Branża
PEKAO i PKO BP	0.345	13.6% i 16.0%	bankowa i bankowa
PEKAO i BZ WBK	0.244	13.6% i 6.2%	
PEKAO i MBANK	0.288	13.6% i 3.6%	
PEKAO i HANDLOWY	0.132	13.6% i 2%	
PEKAO i KGHM	0.247	13.6% i 8.6%	bankowa i wydobywcza
PEKAO i PKN ORLEN	0.211	13.6% i 6.9%	bankowa i paliwowa
PEKAO i ASSECO	0.118	13.6% i 1.9%	bankowa i oprogramowania

Analizując tabelę 35 można potwierdzić, iż podobieństwo czynników lokalnych jest związane zarówno z podobieństwem wielkości spółek, jak również z podobieństwem branż. Jednakże w naszym przypadku wpływ podobieństwa wielkości spółek jest bardziej istotny niż wpływ podobieństwa branży. Oznacza to przewagę innych lokalnych czynników niż różnorodność branży, np. czynników związanych z indywidualnymi strategiami inwestorów.

Do danych dopasowano funkcje Kopuły omówione wcześniej w rozdziale 3.2, zgodnie z procedurą omówioną w rozdziale 3.4.4. Zastosowano tylko te Kopuły, które posiadają indeks dolnej zależności ogonowej, ponieważ analizujemy dane minimalne, dla których stwierdzono występowanie takich zależności. W tabeli 36 pokazano indeksy dolnej zależności ogonowej wyliczone za pomocą Kopuły Gumbela oraz odwróconej Kopuły Claytona [83].

Tabela 36. Indeksy dolnej zależności ogonowej  $\lambda_l$ .

Zmienna	Kopuła Claytona	Odw. Kopuła Gumbela
	$\lambda_l$	$\lambda_l$
PEKAO i PKO BP	0.518	0.426
PEKAO i BZ WBK	0.343	0.312
PEKAO i MBANK	0.425	0.362
PEKAO i HANDLOWY	0.102	0.174
PEKAO i KGHM	0.347	0.314
PEKAO i PKN ORLEN	0.273	0.272
PEKAO i ASSECO	0.074	0.156

Empiryczny indeks dolnej zależności ogonowej wyliczono według wzorów (75) dla  $\alpha = 0.025$ . Podobnie wyliczono indeks dolnej zależności ogonowej, pod warunkiem wystąpienia sygnału anty-korelacji, czyli biorąc pod uwagę jedynie dane, dla których przynajmniej dla jednej ze spółek spełniony jest warunek anty-korelacji  $H \leq H_{\text{threshold}} = 0.45$ . Szczegóły oraz dyskusję ujęto w [83] - natomiast w dysertacji wyniki przedstawiono w tabeli 37.

Tabela 37. Empiryczny indeks dolnej zależności ogonowej, dla wszystkich danych -  $\lambda_l$  oraz dla danych - z warunkiem anty-korelacji -  $\lambda_l(H)$ .

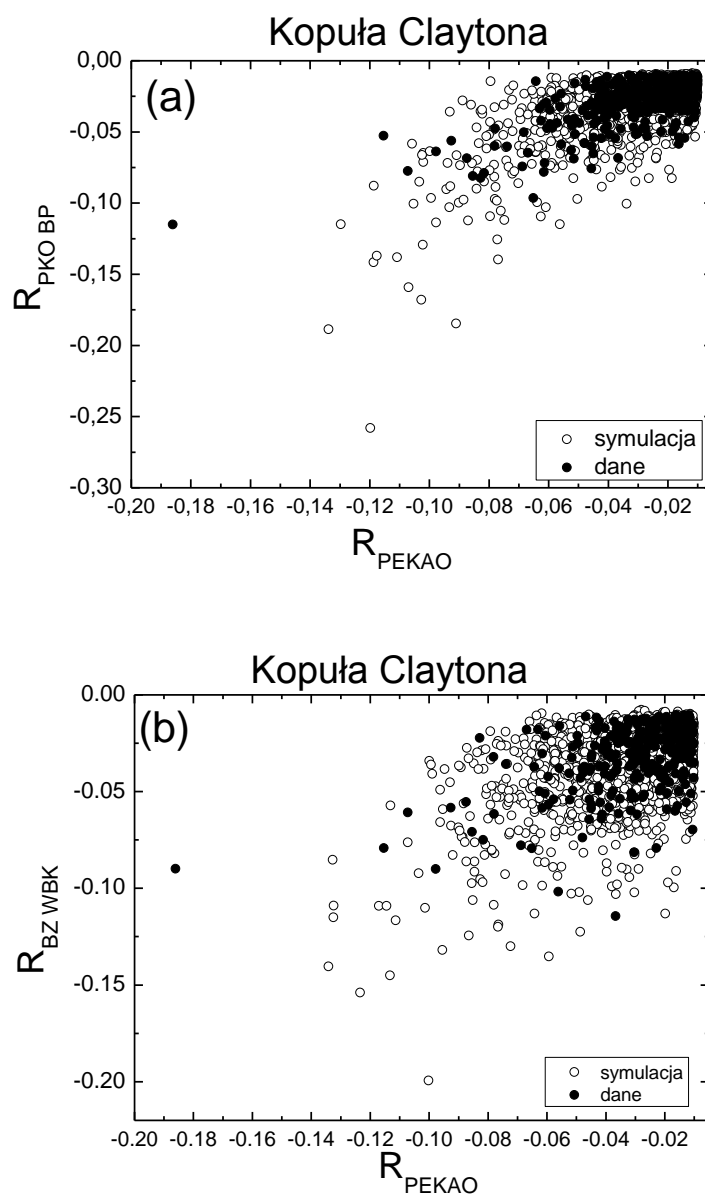
Zmienna	$\lambda_l$	$\lambda_l(H)$
PEKAO i PKO BP	0.482	0.554
PEKAO i BZ WBK	0.325	0.392
PEKAO i MBANK	0.410	0.480
PEKAO i HANDLOWY	0.163	0.181
PEKAO i KGHM	0.135	0.159
PEKAO i PKN ORLEN	0.387	0.391
PEKAO i ASSECO	0.335	0.372

Analizując tabelę 37 można stwierdzić, że sygnał anty-korelacji, wyznaczony w wykorzystaniem wykładnika Hursta oraz DFA, zwiększa ryzyko wystąpienia jednoczesnych silnych spadków dla obu papierów wartościowych będących w portfelu inwestycyjnym. Jest to zgodne z wynikami

wcześniejszych badań [2, 24, 25]. Analizując tabelę 37 można zauważyć ponadto, że malejąca korelacja nie zawsze odzwierciedla się w malejącej dolnej zależności ogonowej [83]. Także podczas badań maksymalnych wezbrań sztormowych dokonano takiej obserwacji - tabela 31, co wynika z różnej auto-korelacji poszczególnych szeregów czasowych. W przypadku danych finansowych obserwujemy zmienną auto-korelację również różniącą się dla poszczególnych spółek. Warto w tym miejscu zwrócić uwagę na czynnik globalny będący kolektywną strategią inwestorów, którzy w przypadku wystąpienia dużych spadków sprzedają swoje akcje powodując kolejne spadki. Można wywnioskować, że czynnik ten nie zawsze współgra z pozostałymi czynnikami globalnymi, ogólnie wprowadzającymi korelację. Okazuje się, że czynniki związane z kolektywną lub indywidualną strategią inwestorów posiadają znaczne auto-korelacje, gdyż „panika” na rynku może trwać wiele dni. Dlatego tego typu czynnik globalny lub lokalny może powodować z jednej strony pojawienie się sygnału anty-korelacji, natomiast z drugiej strony wystąpienie ekstremalnych ujemnych stóp zwrotu.

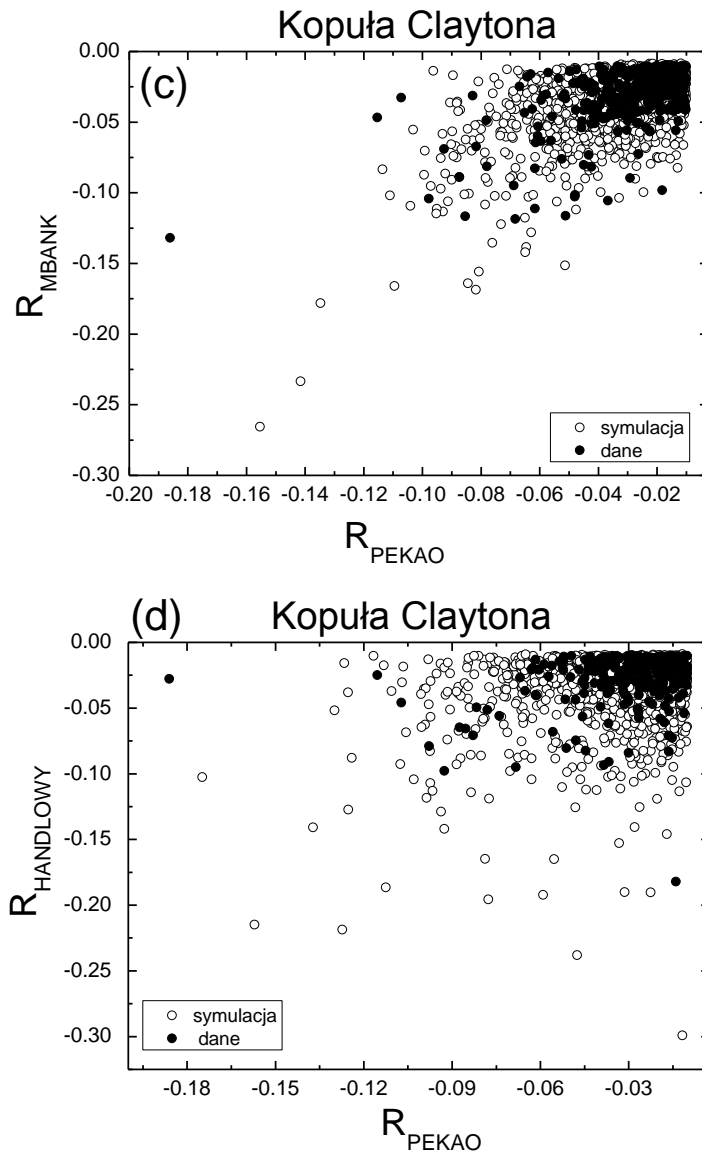
Zauważmy, porównując tabelę 36 oraz 37, że wykładnik Hursta można wykorzystać w procedurze wyboru odpowiedniej funkcji Kopuły [83]. W przypadku par banków posiadających podobne czynniki lokalne odnoszące się do sektora bankowego sygnał anty-korelacji pozwala na wybór pomiędzy Kopułą Clayтона oraz odwróconą Kopułą Gumbela, z których obie dobrze modelują dane. Porównując tabele 30 i 31 z tabelami 36 oraz 37 możemy zauważyć, że model ten jest analogiczny do modelu oceny ryzyka wystąpienia maksymalnych wezbrań sztormowych, z mającymi miejsce istotnymi różnicami. W przypadku danych morskich występują większe wartości korelacji i zależności ogonowej niż w przypadku danych finansowych. W przypadku wezbrań sztormowych badamy globalny sygnał anty-korelacji, a w przypadku danych finansowych sygnał lokalny. Ostatecznie w pierwszym przypadku badamy dane maksymalne, a w drugim minimalne. Wracając do tabeli 36 oraz 37 można zauważyć, że w przypadku pary zawierającej bank oraz spółkę niebędącą bankiem, wymienione tam funkcje Kopuły gorzej modelują wartości ekstremalne. Jednak można ogólnie powiedzieć, że odwrócona Kopuła Gumela prezentuje lepszy model. Na rysunku 39 przedstawiono dane modelowane przez Kopułę Clayтона dla banków, natomiast na rysunku 40 przez odwróconą Kopułę Gumbela dla pary bank - spółka niebędąca bankiem.

W pracy [83] omówiono pewien istotny aspekt uniwersalny dla wszystkich badanych par. Brak anty-korelacji ( $H > 0.45$ ) dla obu spółek należących do portfela powoduje, iż empiryczny indeks dolnej zależności ogonowej jest równy zero, co pozwala użyć do modelowania ryzyka funkcji Kopuły nieposiadających zależności ogonowej np. Kopuły Gaussa, która ma szerokie aczkolwiek dyskusyjne [1, 2] zastosowanie w finansach. Niestety, wymieniony wyżej warunek jest spełniony tylko dla 10-20% czasu transakcji.

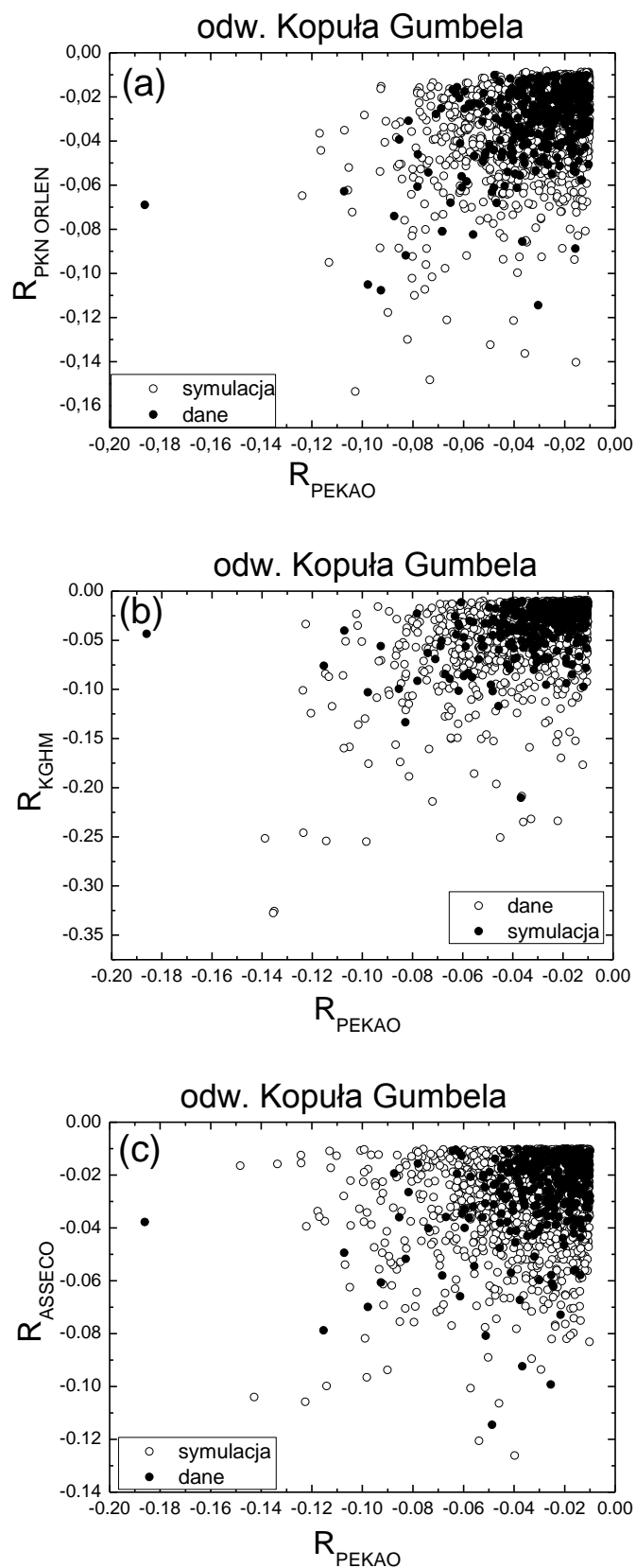


Rysunek 39. Dane empiryczne dla par banków oraz dane zasymulowane w przestrzeni stóp zwrotu -  
Kopuła Claytona: (a) - PEKAO i PKO BP, (b) - PEKAO i BZ WBK.





Rysunek 39 – kont. Dane empiryczne dla par banków oraz dane zasymulowane w przestrzeni stóp zwrotu - Kopuła Claytona: (c) - PEKAO i MBANK, (d) - PEKAO i Handlowy.



Rysunek 40. Dane empiryczne dla par bank i spółka nie będąca bankiem oraz dane zasymulowane w przestrzeni stóp zwrotu - odwrócona Kopuła Gumbela: (a) - PEKAO i PKN ORLEN, (b) - PEKAO i KGHM, (c) - PEKAO i ASSECO.

Tabela 38. Parametr Kopuły Claytona oraz odwróconej Kopuły Gumbela.

Zmienna	Kopuła Claytona	Odw. Kopuła Gumbela
	$\theta$	$\theta$
PEKAO i PKO BP	1.055	1.528
PEKAO i BZ WBK	0.647	1.324
PEKAO i MBANK	0.811	1.405
PEKAO i HANDLOWY	0.303	1.152
PEKAO i KGHM	0.655	1.328
PEKAO i PKN ORLEN	0.534	1.267
PEKAO i ASSECO	0.266	1.133

W tabeli 38 przedstawiono parametry dopasowanych funkcji Kopuły. Warto w tym miejscu zaznaczyć, że model oceny ryzyka inwestycyjnego dla par papierów wartościowych jest podobny do modelu oceny ryzyka wystąpienia wysokich wezbrań sztormowych w kilku lokalizacjach. Ryzyko zostało określone przez funkcje Kopuły posiadającą dolny indeks zależności ogonowej – Kopułę Claytona lub odwróconą Kopułę Gumbela. Auto-korelacja wyznaczona przy pomocy lokalnego DFA okazuje się przydatnym narzędziem podczas procedury wyboru właściwej funkcji Kopuły, szczególnie dla par dwóch banków. Okazało się ponadto, iż czynnik lokalny związany z indywidualną strategią inwestorów ma stosunkowo duży wpływ na kursy akcji powodując lepszą korelację dla spółek o podobnej wielkości.

## 6. Podsumowanie pracy

W ramach podsumowania wyników pracy warto zwrócić uwagę na mechanizmy wspólne dla wszystkich analizowanych danych. Takim mechanizmem jest wspólny udział procesów czasowych o charakterze zarówno deterministycznym, jak i stochastycznym, na które wpływ mają czynniki globalne, wprowadzające korelacje dla wszystkich danych, oraz czynniki lokalne, wprowadzające korelacje dla par bliskich danych oraz losowość dla par odległych danych.

W przypadku koncentracji jonów pochodzenia morskiego oraz lądowego w rdzeniach arktycznych, globalnym czynnikiem jest oddziaływanie układów cyklonicznych odpowiedzialnych za transport aerozoli w kierunku południowym. Lokalne czynniki to przede wszystkim intensywność wiatrów wiejących nad obszarami lądowymi i morskimi oraz intensywność wiatrów wiejących nad różnymi obszarami lądowymi, z których pochodzą indywidualne jony. Dla jonów pochodzenia lądowego stwierdzono silną auto-korelację związaną prawdopodobnie z akumulacją pierwiastków w obszarze lądowym. Dla jonów pochodzenia morskiego stwierdzono zmienną auto-korelację analogiczną jak dla danych finansowych.

W przypadku maksymalnych wezbrań sztormowych czynnikiem globalnym są północno - zachodnie wiatry oraz trwające przez kilkanaście godzin atmosferyczne fronty niskiego ciśnienia. Czynniki lokalne związane są z różnymi lokalnymi kierunkami ruchu frontów niskiego ciśnienia oraz różniącym się lokalnie kształtem linii brzegowej. Różna auto-korelacja dla danych zarejestrowanych w różnych lokalizacjach świadczy o auto-korelacji i pewnej przewidywalności niektórych czynników lokalnych.

W przypadku danych finansowych czynniki globalne to między innymi koniunktura oraz ogólny stan rynku związany z kolektywnym zachowaniem inwestorów. Czynniki lokalne są związane z poszczególnymi spółkami oraz sektorami jak również indywidualnymi strategiami inwestorów. Kolektywne zachowanie inwestorów odpowiada za globalne auto-korelacje [19 - 23], natomiast ich indywidualne strategie za lokalne auto-korelacje [2, 24, 25].

Kolejnym ważnym wynikiem badań jest analogia pomiędzy wpływem dynamiki wykładnika Hursta na dane finansowe oraz koncentracja jonów pochodzenia morskiego (sodu i magnezu) w rdzeniach arktycznych. W przypadku danych finansowych spadek wartości wykładnika Hursta jest

związany ze zbliżającą się zmianą trendu lub silnymi spadkami kursów akcji. W tabeli 37 pokazano, że duże spadki kursów akcji są bardziej prawdopodobne w punkcie, w którym wykryto sygnał anty-korelacji. Zjawisko to tłumaczą wieloletnie badania nad analogią pomiędzy układem finansowym i dynamicznym układem złożonym [2]. Fakt ten można też wytłumaczyć zakładając anty-korelację lokalnych lub globalnych czynników związanych z indywidualną lub kolektywną strategią inwestorów. Analogiczne zjawisko stwierdzono dla koncentracji jonów pochodzenia morskiego w rdzeniach lodowych. Analizując rysunek 21 można stwierdzić występowanie sygnału anty-korelacji  $H < 0.5$  dla ostatnich 0.8-1.5 tys. lat. Dla obu jonów można było stwierdzić, w tym okresie, wartość wykładnika Hursta mniejszą od wartości progowej  $H_{threshold} = 0.47$ . Po zarejestrowaniu tego sygnału anty-korelacji, można było stwierdzić wyższe koncentracje obu jonów niż we wcześniejszym okresie. Ponieważ anty-korelacja nie występowała dla innych jonów pochodzenia lądowego, należy wywnioskować, że jest ona związana z lokalnym czynnikiem, którym jest intensywność wiatrów wiejących nad obszarami morskimi. Dlatego za wyższą koncentrację jonów musi odpowiadać pewna zmiana modelu cyrkulacji tych wiatrów, która miała miejsce około 1.5 tys. lat temu, analogiczna do zmiany strategii inwestorów, która może wpływać na silne spadki kursów akcji. Wykładnik Hursta wyliczony dla koncentracji jonów pochodzenia morskiego wykazywał wyższe wartości niż wyliczony dla kursów akcji, gdzie należało stosować niższą wartość progową  $H_{threshold} = 0.45$ . Jest to spowodowane prawdopodobnie większą zmiennością i nieprzewidywalnością wartości wykładnika Hursta wyliczonego dla danych finansowych w porównaniu z wykładnikiem Hursta wyliczonym dla danych fizycznych.

Istotnym wynikiem badań jest również wspólny model oceny ryzyka zastosowany do analizy danych finansowych oraz geofizycznych. W przypadku danych finansowych ocenia się ryzyko wystąpienia jednoczesnych ekstremalnych spadków kursów akcji dwóch spółek należących do portfela inwestycyjnego, natomiast w przypadku geofizycznym ryzyko wystąpienia ekstremalnych wezbrań sztormowych w dwóch lokalizacjach. Dane finansowe składające się z pary papierów wartościowych dwóch banków notowanych na GPW w Warszawie najlepiej modelowała Kopuła Gumbela oraz odwrócona Kopuła Claytona, a analiza auto-korelacji danych okazała się pomocna w wyborze między tymi dwoma funkcjami Kopuły. W przypadku maksymalnych wezbrań sztormowych dane najlepiej modelowała odwrócona Kopuła Gumbela oraz Kopuła Claytona. Podobnie jak w przypadku danych finansowych badania auto-korelacji okazały się przydatne w procedurze wyboru funkcji Kopuły. Parametry funkcji Kopuły dla danych finansowych i geofizycznych można porównać zestawiając tabele 32 oraz 38. W przypadku danych geofizycznych notujemy ogólnie wyższe wartości korelacji, jednak istnieją pary danych o podobnych korelacjach, a co tym idzie podobnych parametrach funkcji Kopuły. Takimi parami są:

- maksymalne wezbrania sztormowe w Wismar i Kołobrzegu oraz ekstremalne ujemne stopy zwrotu dla PEKAO i BZ WBK i PEKAO i KGHM,
- maksymalne wezbrania sztormowe w Warnemunde i Kołobrzegu oraz ekstremalne ujemne stopy zwrotu dla PEKAO i MBANK.

W pracy podjęto również próbę porównania modelu ekstremalnie niskich stóp zwrotu z maksymalnymi koncentracjami jonów pochodzenia morskiego (sodu i magnezu) w rdzeniach lodowych - rozdział 5.1.4, rysunek 28. Motywacją do tych badań była podobna dynamika wykładnika Hursta dla obu rodzajów danych. Niestety, dla maksymalnych koncentracji jonów korelacja była bardzo silna, a parametr Kopuły Claytona dużo wyższy niż te zarejestrowane dla danych finansowych - tabela 38. Wynika to z faktu, że dane finansowe są w dużej mierze stochastyczne. Dlatego współczynniki korelacji mogą być dla nich niższe niż dla danych fizycznych. Różnice pomiędzy danymi finansowymi oraz fizycznymi są omówione poniżej.

W podsumowaniu stwierdza się występowanie następujących analogii pomiędzy danymi geofizycznymi oraz finansowymi:

- podobna dynamika wykładnika Hursta dla danych finansowych oraz koncentracji jonów pochodzenia morskiego w rdzeniach arktycznych,
- analogiczny model, według którego badane szeregi czasowe są częściowo stochastyczne i częściowo deterministyczne, a ich mniejsze lub większe korelacje zawdzięczają zarówno czynnikom lokalnym jak i globalnym,
- wspólny model oceny ryzyka wystąpienia maksymalnych spadków kursów papierów wartościowych oraz wystąpienia wysokich wezbrań sztormowych.

Ogólnie analogie są oparte na fakcie, iż omawiane układy zarówno fizyczne jak i finansowe są na tyle złożone, że ich deterministyczny opis nie jest możliwy, z drugiej strony nie są one w pełni stochastyczne. Poszukiwaniem analogii pomiędzy danymi fizycznymi oraz finansowymi od wielu lat zajmuje się ekonofizyka, a tworzenie modeli oceny ryzyka jest w ostatnich latach dużym wyzwaniem w różnych dziedzinach nauki.

W pracy omówiono model oceny ryzyka, który może mieć zastosowanie zarówno do badania danych geofizycznych jak i finansowych. W przypadku danych finansowych badaniom podlegały dzienne kursy akcji oraz dzienne stopy zwrotu; okazało się, że wykładnik Hursta wyliczony z użyciem lokalnej DFA może okazać się przydatny w procedurze oceny ryzyka spadku wartości obu portfela. Rozwój tego typu modeli będzie miał coraz większe zastosowanie w badaniu danych finansowych oraz różnego rodzaju zjawisk fizycznych. Na koniec zostanie omówiona propozycja rozwoju takiego

modelu do badania danych finansowych. Należy jednak pamiętać, że możemy podobne modele zastosować do badania danych fizycznych.

Ogólnie zachowanie się kursów papierów wartościowych jest inne w krótkiej i długiej skali czasowej [2]. Dlatego osobno powinno się badać auto-korelacje oraz korelacje pomiędzy danymi finansowymi (cenami akcji) w krótkiej skali czasowej - minuty, sekundy, każda transakcja, oraz długiej skali czasowej - dzienne transakcje. Poniżej zostanie przedstawiona w kilku krokach procedura, którą można użyć do badania danych finansowych, kursów papierów wartościowych  $P_i(t)$ , gdzie indeks  $i$  numeruje kolejne spółki. Procedura została opracowana na podstawie badań przedstawionych w obecnej pracy.

1. Transformacja danych - wyliczenie stóp zwrotu  $R_i(t) = \frac{P_i(t+1)-P_i(t)}{P_i(t)}$ .
2. Określenie ekstremalnych stóp zwrotu (ekstremalnej straty)  $R_{\alpha,i}(t) < \alpha$  (np.  $\alpha < 0.02, \dots 0.05$ );
3. dopasowanie brzegowych rozkładów prawdopodobieństwa  $f_i(x_i)$  wartości ekstremalnych  $R_{\alpha,i}(t) < \alpha$ .
4. Transformacja danych do przestrzeni  $[0,1]$  z użyciem dystrybuanty  $F_i(x_i)$ .
5. Utworzenie  $n$  - akcyjnego portfela, przejście do przestrzeni  $[0,1]^n$ , w przypadku dwuskładnikowych portfeli (omówionego w pracy) do przestrzeni  $[0,1]^2$ .
6. Dopasowanie odpowiednich funkcji Kopuły.
7. Badanie zależności ogonowych funkcji Kopuły.
8. Obliczenie dla badanych kursów akcji  $P_i(t)$  wykładnik Hursta z użyciem DFA w celu zbadania warunku anty-korelacji  $H < H_{threshold}$  (np.  $H_{threshold} = 0.45$ ) jako warunku na ujemną auto-korelację i możliwe załamanie na rynku.
9. Wykorzystanie wyników z pkt. 8 do analizy zależności ogonowych.
10. Ewaluacja modelu oceny ryzyka typu VaR, w którym:
  - a) ryzyko jest określone przy pomocy Funkcji Kopuły;
  - b) sygnał anty-korelacji jest użyty w procedurze wyboru funkcji Kopuły.
11. Zastosowanie modelu oraz odpowiednio dopasowanych funkcji Kopuły od faktycznej oceny ryzyka oraz konstrukcji optymalnego portfela inwestycyjnego (minimalizującego ryzyko przy ustalonym przewidywanym zysku lub maksymalizującego przewidywany zysk przy ustalonym poziomie ryzyka).

Powołując się na analogie, w podobny sposób można budować modele oceny prawdopodobieństwa wystąpienia ekstremalnych danych dla układu fizycznego lub geofizycznego.

Warto jednak się zastanowić, w jakim zakresie można szukać analogii pomiędzy układami fizycznymi oraz ekonomicznymi. Należy pamiętać, że dla danych finansowych nie ma analogii do stałych w czasie i przestrzeni praw fizyki [84] z których wynikają prawa zachowania. Dlatego warto prowadzić dalsze badania, aby odpowiedzieć na pytanie, do jakiego stopnia analogia pomiędzy fizyką i ekonomią może być użyta do przybliżenia świata realnego.



## Literatura

- [1] U. Cherubini, E. Luciano, W. Vecchiato, *Copula Methods in finance*, Wiley finance (2012).
- [2] K. Domino, J. Dajka, *Fizyka w ekonomii metody i modele*, Wydawnictwo UŚ (2014).
- [3] S. Havlin, D. Ben-Avraham, *Diffusion in disordered media*, Advances in Physics, 51 (2002) 187.
- [4] B. B. Mandelbrot, *Fractals: Form, Chance and Dimension* (San Francisco: Freeman) 1977.
- [5] V. Andreychouk, T. Błachowicz, K. Domino, *Fractal dimensions of cave for exemplary gypsum cave-mazes of Western Ukraine*, Landform Analysis, Vol. 22: 3–8 (2013).
- [6] V. Andreychouk, T. Błachowicz, K. Domino, *Fractal dimensions of gypsum cave-mazes of Western Ukraine*, Speleology and Karstology, Vol. 11 (2013) 40-47.
- [7] S. Alexander, S., R. Orbach, *Density of states on fractals: "fractons."*, J. Phys. Lett., Paris, **43**, L625 (1982).
- [8] C. Kittel, *Introduction to Solid State Physics*, New York: Wiley (1971).
- [9] T. Błachowicz, V. Andreychouk, K. Domino, *Quantitative estimation of structural complexity using random walking and anomalous diffusion formalisms for exemplary gypsum cave-mazes of Western Ukraine*, Mathematical Geosciences (In Review).
- [10] B. Mandelbrot and J. W. van Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM Rev. 10, pp. 422-437 (1968).
- [11] S. Havlin and D. Ben-Avraham, *Fractal dimensionality of polymer chains*, J. Phys. A15, L311-L316 (1982).
- [12] G.L. Vasconcelos, *A guided walk down Wall Street: an Introduction to Econophysics*, Brazilian Journal of Physics 34 (2004) 1039.

- [13] M. S. Taqqu, V. Teverovsky, W. Willinger, *Estimators for long-range dependence: an empirical study*, *Fractals* 3 (1995) 785.
- [14] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-Term Storage: An Experimental Study*, Constable, London (1965).
- [15] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter and K. Voigt, *Comment on Delocalization in the 1D Anderson model with long-range correlated disorder*, *Phys. Rev. Lett.* 85 (2000) 3736.
- [16] N. Makarava, S. Benmehdi, M. Holschneider, *Bayesian estimation of self-similarity exponent*, *PHYSICAL REVIEW E* 84, 021109 (2011).
- [17] B.B. Mandelbrot, *The variation of certain speculative prices*, *Journal of Business* 36 (1963) 394.
- [18] B.B. Mandelbrot, *Fractals and Scaling in Finance*, Springer-Verlag, New York, 1997.
- [19] D. Grech, G. Pamula, *The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market*, *Physica A* 387 (2008) 4299.
- [20] L. Czarnecki, D. Grech, G. Pamula, *Comparison study of global and local approaches describing critical phenomena on the Polish stock exchange market*, *Physica A* 387 (2008) 6801–6811.
- [21] R.L. Costa, G.L. Vasconcelos, *Long-range correlations and nonstationarity in the Brazilian stock market*, *Physica A* 329 (2003) 231.
- [22] D. Grech, Z. Mazur, *Can one make any crash prediction in finance using the local Hurst exponent idea?*, *Physica A* 336 (2004) 133.
- [23] N. Vandewalle, M. Ausloos, *Coherent and random sequences in financial fluctuations*, *Physica A* 246 (1997) 454.
- [24] K. Domino, *The use of the Hurst exponent to predict changes in trends on the Warsaw Stock Exchange*, *Physica A* 390 (2011) 98.
- [25] K. Domino, *The use of the Hurst exponent to investigate the global maximum of the Warsaw Stock Exchange WIG20 index*, *Physica A* 391 (2012) 156.

- [26] J.M.W. Tadion, *Deciphering the Market: Principles of Chart Reading and Trading Stocks, Commodities, and Currencies*, Wiley, New York, 1996.
- [27] D. Sornette, A. Johansen, J.P. Bouchaud, *Stock Market Crashes, Precursors and Replicas*, Journal de Physique I (France) 6 (1996) 167.
- [28] J.A. Feigenbaum, P.G.O. Freund, *Discrete scale invariance in stock markets before crashes*, International Journal of Modern Physics B 10 (1996) 3737.
- [29] D. Sornette, A. Johansen, *Large financial crashes*, Physica A 245 (1997) 411.
- [30] N. Vandewalle, P. Boveroux, A. Minguet, M. Ausloos, *The crash of October 1987 seen as a phase transition: amplitude and universality*, Physica A 255 (1998) 201.
- [31] N. Vandewalle, M. Ausloos, Ph. Boveroux, A. Minguet, *How the financial crash of October 1997 could have been predicted*, The European Physical Journal B 4 (1998) 139.
- [32] N. Vandewalle, M. Ausloos, Ph. Boveroux, A. Minguet, *Visualizing the log-periodic pattern before crashes*, The European Physical Journal B 9 (1999) 355.
- [33] [www.gpw.pl](http://www.gpw.pl)
- [34] P. Best, *Implementing Value at Risk*, Wiley Series in Financial Engineering (1999).
- [35] P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.), McGraw-Hill, ISBN 978-0-07-146495-6 (2006).
- [36] L. Kristoufek, *Local Scaling Properties and Market Turning Points at Prague Stock Exchange*, Acta Physica Polonica B, Vol. 41 (2010) No. 6.
- [37] P. Bak, M. Paczuski, M. Shubik, *Price variation in a stock market with many agents*, Physica A 246, 430-453 (1997).
- [38] N. Kruszewska, P. Weber, A. Gadomski, K. Domino, *A method of mechanical control of structure-property relationship in grain-containing material systems*, Acta Physica Polonica B, Vol. 44 (2013), No 5.
- [39] M. Gligor, M. Ignat, *Econophysics: a New Field for Statistical Physics*, Interdisciplinary Science Reviews, VOL. 26, NO. 4 (2001).

- [40] A. Sklar, *Fonctions de Repartition a N Dimensions et Leurs Marges*, Publ. Inst. Statist. Univ. Paris. 8, 229 (1959).
- [41] R. J. Scherrer, A. A. Berlind, Q. Mao, C. K. McBride, *From finance to cosmology: the copula of large-scale structure*, arXiv:0909.5187v2.
- [42] R.-G. Cong, M. Brady, *The Interdependence between Rainfall and Temperature: Copula Analyses*, The Scientific World Journal (2012).
- [43] C. Schölzel, P. Friederichs, *Multivariate non-normally distributed random variables in climate research – introduction to the copula approach*, Nonlinear Processes in Geophysics 15, 761 (2008).
- [44] D. B. Pougaza, A.M. Djafari, J.-F. Bercher, *Using the Notion of Copula in Tomography*, arXiv:0812.1316v1.
- [45] A. Onken, S. Grünewälder, M. H. Munk, K. Obermayer, *Analyzing Short-Term Noise Dependencies of Spike-Counts in Macaque Prefrontal Cortex Using Copulas and the Flashlight Transformation*, in Aertsen, Ad, PLoS Computational Biology 5, (2009).
- [46] H. M. N. D. Bandara, A. P. Jayasumana, *On Characteristics and Modeling of P2P Resources with Correlated Static and Dynamic Attributes*, IEEE GLOBECOM: 1–6 (Dec. 2011).
- [47] S. Demarta, A. J. McNeil, *The t-Copula and Related Copulas*, Department of Mathematics Federal Institute of Technology (2004).
- [48] A.J. McNeil, J. Neslehova, *Multivariate Archimedean Copulas*, Maxwell Institute Edinburgh & ETH Zurich (2007).
- [49] H. Joe, J.J. Xu, *The Estimation Method of Inference Functions for Margins of Multivariate Models*. Dept. of Statistics University of British Columbia, Tech. Rep. 166.
- [50] S. Demarta, A. J. McNeil, *The t Copula and Related Copulas*, International Statistical Review 73, 111–129 (2005).
- [51] R. Mashal, M. Naldi, A. Zeevi, *On the dependence of equity and asset returns*, RISK, 16, 83–87 (2003).
- [52] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York (1958).

- [53] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York (1978).
- [54] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York (1983).
- [55] M. Fréchet, Sur la loi de probabilité de l'écart maximum, *Annales de la Société Polonaise de Mathématique*, Cracovie 6, 93 (1927).
- [56] R. A. Fisher, L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Philos. Soc.* 24, 180 (1928).
- [57] E. J. Gumbel, Les valeurs extremes des distributions statistiques, *Annales de l'Institut Henri Poincaré*, 4, 115 (1935).
- [58] E. J. Gumbel, The return period of flood flows, *Ann. Math. Statistics*, 12, 163 (1941).
- [59] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, *Extreme value statistics in records with long-term persistence*, *Phys. Rev. E* 73, 016130 (2006).
- [60] P.J.A Caton, D.S. Mapa, *Time-Varying Conditional Johnson Su Density In Value-at-Risk (VaR) Methodology*, UPSS Working Paper No. 2012-01 (2012).
- [61] F. Ashkar, F. Aucoin, Discriminating between the Lognormal and the Log-Logistic Distributions for Hydrological Frequency Analysis, *J. Hydrol. Eng.*, 17(1), 160 (2012).
- [62] Z. A. Zakaria, A. Shabri, U. Nadiah, Estimation of the generalized logistic distribution of extreme events using partial L-moments, *Hydrological Sciences Journal*, vol. 57, Issue 3, (2012).
- [63] P. Jaworski, F. Durante, W. Hardle, T. Rychlik, *Copula theory and its Applications*, Springer-Verlag Berlin Heidelberg (2010).
- [64] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, *Statistics of return intervals in long-term correlated records*, *Physical Review E* 75, 011128 (2007).
- [65] P.A. Mayewski, M.S. Twickler, S.I. Whitlow, L.D. Meeker, Q. Yang, J. Thomas, K. Kreutz, P.M. Grootes, D.L. Morse, E.J. Steig, E.D. Waddington, E. Saltzman, P-Y. Whung, K. Taylor, *Climate Change During the Last Deglaciation in Antarctica*, *Science* 272: 1636-1638 (1996)

- [66] E.J. Steig, E.J. Brook, J.W.C. White, C.M. Sucher, M.L. Bender, S.J. Lehman, D.L. Morse, E.D. Waddington, G.D. Clow, Synchronous climate changes in Antarctica and the North Atlantic, *Science* 282 (5386), 92-95 (1998).
- [67] E.J. Steig, D.L. Morse, E.D. Waddington, M. Stuiver, P.M. Grootes, P.A. Mayewski, M.S. Twickler, S.I. Whitlow, Wisconsinan and Holocene climate history from an ice core at Taylor Dome, western Ross Embayment, Antarctica, *Geografiska Annale*, 82 A (2–3): 213–235 (2000)
- [68] P.M. Grootes, E.J. Steig, M. Stuiver, E.D. Waddington, D.L. Morse DL., GISP2 - Taylor Dome oxygen isotope ratios. *Quat. Res.* (in review) (1999).
- [69] E.J. Steig, C.P. Hart, J.W.C. White, W.L. Cunningham, M.D. Davis, E.S. Saltzman, Changes in climate, ocean and ice sheet conditions in the Ross, Embayment at 6 ka, *Annals of Glaciology* 27, 305-310 (1998).
- [70] <http://tools.wmflabs.org/geohack>
- [71] J.C. Stager, P.A. Mayewski, Abrupt early to mid-Holocene climate transition registered at the equator and the poles, *Science* 276: 1834-1836 (1997)
- [72] <http://www.iceandclimate.nbi.ku.dk>
- [73] G. M. Raisbeck, F. Yiou, Production of long-lived cosmogenic nuclei and their applications, *Nucl. Instrum. Methods, Sect. B*, 5, 91-99, 1984.
- [74] D. Lal, Theoretically Expected Variations in the Terrestrial Cosmic - Ray Production Rates of Isotopes. In: Castagnoli, G.C. (Ed.), *Theoretically Expected Variations in the Terrestrial Cosmic-Ray Production Rates of Isotopes*, (1988) North-Holland, Amsterdam, XCV, pp. 215–233.
- [75] E.J. Steig, Beryllium-10 in the Taylor Dome ice core: applications to Antarctic glaciology and paleoclimatology (1996) *Antarctic Journal of the United States* 31: 71-72.
- [76] Raisbeck GM, Yiou F, 10Be in polar ice and atmosphere, (1985) *Ann Glaciol* 7:138–140.
- [77] SOURCES OF CONTINENTAL DUST OVER ANTARCTICA DURING THE LAST GLACIAL CYCLE (1992), *Journal of Atmospheric Chemistry* 14: 233-244
- [78] Magazine - Water Condition & purification, January (2005).
- [79] K. Kawamura, F. Parrenin, L. Lisiecki, R. Uemura, F. Vimeux, J.P. Severinghaus, M. A. Hutterli, T. Nakazawa, S. Aoki, J. Jouzel, M. E. Raymo, K. Matsumoto, H. Nakata, H. Motoyama, S. Fujita, K. Goto-

## Literatura

Azuma, Y. Fujii, O. Watanabe, Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature*, Vol. 448, pp. 912-916 (2007).

[80] K. Domino, T. Blachowicz, M. Ciupak, *The use of copula functions for predictive analysis of correlations between extreme storm tides*, *Physica A* 413 (2014) 489.

[81] M. Ciupak, K. Rokciński, *Zastosowanie metody kopuli do dwuwymiarowej analizy wezbrań sztormowych w profilach wodowskazowych Świnoujścia i Kołobrzegu*, *Zeszyty naukowe akademii marynarki wojennej*, 4 (2011) 15-34.

[82] [www.google.pl/maps](http://www.google.pl/maps).

[83] K. Domino, T. Błachowicz, *The use of copula functions for modeling the risk of investment in shares traded on the Warsaw Stock Exchange*, *Physica A* 413 (2014) 77.

[84] M.F.M Osborne, *The Stock Market and Finance From a Physicist's Viewpoint*, Minneapolis: Crossgar Press, cop. (1977).

## Dodatki

Dodatki o charakterze numerycznym, to skrypty napisane w języku Python 2.7, które zostały użyte do analizy danych. Opisy oraz odnośniki do skryptów znajdują się w rozdziałach 5.1.2 oraz 5.1.3.

### Dodatek A - transformacja danych, utworzenie szeregów skumulowanych

```
import numpy
from scipy import polyfit
from numpy import zeros, floor, log10, log, mean, array
import matplotlib.pyplot as plt
# szereg skumulowany dla Na
nazwa = "daneNaMg"
ls = []
dat = []
# wczytuje dane
c = -1
for line in open(str(nazwa)+".txt", "r").readlines():
    c = c + 1
    if c > 0:
        a = 0
        for value in line.split(","):
            a = a + 1
            if a == 2:
                value = float(value)
                ls.append(value)
# wczytuje date
dat = []
c = -1
for line in open(str(nazwa)+".txt", "r").readlines():
    c = c + 1
    if c > 0:
        a = 0
        for value in line.split(","):
            a = a + 1
            if a == 1:
                value = float(value)
                dat.append(value)

# wylicza szereg skumulowany
lw = ls - mean(ls)

l = []
l.append(lw[0])
for j in range(1, len(lw)):
    l.append(l[j-1]+lw[j])

# plik wyjściowy
out_file = open("danemodNa.txt", "w")
```



```

out_file.write("date, delta, \n")
for i in range(0, len(l)):
    out_file.write(str(dat[i])+", "+str(l[i])+"\n")
out_file.close()

#szereg skumulowany dla ["Mg", "Cl", "Ca", "K"]
name = ["Mg", "Cl", "Ca", "K"]
# petla rozne pliki - pozostale operacje jw.
for f in range(0, len(name)):
    nazwa = "daneNa"+str(name[f])
    ls = []
    dat = []

    c = -1
    for line in open (str(nazwa)+".txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 3:
                    value = float(value)
                    ls.append(value)
    dat = []
    c = -1
    for line in open (str(nazwa)+".txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 1:
                    value = float(value)
                    dat.append(value)

    lw = ls - mean(ls)

    l = []
    l.append(lw[0])
    for j in range(1, len(lw)):
        l.append(l[j-1]+lw[j])

    out_file = open("danemod"+str(name[f])+".txt", "w")
    out_file.write("date, delta, \n")
    for i in range(0, len(l)):
        out_file.write(str(dat[i])+", "+str(l[i])+"\n")
    out_file.close()

```

## Dodatek B - wyliczenie wykładnika Hursta z użyciem DFA

```
import numpy
from scipy import polyfit
from scipy import stats
from numpy import zeros, floor, log10, log, mean, array
import matplotlib
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt

# wyliczenia w petli dla roznych plikow
naz = ["Na", "K", "Cl", "Mg", "Ca"]
for ij in range(0, len(naz)):
    nazwa = "danemod"+str(naz[ij])
    # maksymalna ilosc pod serii
    limit = 100
    ls = []
    c = -1
    for line in open (str(nazwa)+".txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 2:
                    value = float(value)
                    ls.append(value)
    dat = []

    c = -1
    for line in open (str(nazwa)+".txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 1:
                    value = float(value)
                    dat.append(value)

# wylicza wariancje beztrendowa
def Fcalc(P1, P2):
    P2 = array(P2)
    N = len(P1)
    (a,b) = polyfit(P2,P1,1)
    P_1 = a*P2 + b
    F2 = sum((P1-P_1)**2)/N
    return F2

# wylicza wykladnik Hursta
```

```

def hurst(P, p2):
    N = len(P)
    row = []
    series = []
    i = 1
    while i <= limit:
        tau = N/i
        if tau * i != N:
            tau = N/i + 1
        series.append(log(tau))
        j = 1
        wynik = 0
        while j * tau < N:
            wynik = Fcalc(P[(j-1)*tau:j*tau], p2[(j-1)*tau:j*tau]) + wynik
            j = j+1
        wynik = Fcalc(P[N-tau:N], p2[N-tau:N]) + wynik
        wynik = 0.5 * log(wynik / j)
        row.append(wynik)
        i = i+1
    slope, intercept, r_value, p_value, std_err = stats.linregress(series, row)

    return slope, std_err
# zapisuje do pliku wykladnik Hursta
out_file = open(str(naz[ij])+"_hurst.txt", "w")
out_file.write(str("H")+","+str("error")+","+str("date")+"\n")
for i in range(0, len(Is)-1000):
    out_file.write(str(hurst(Is[i:1000+i], dat[i:1000+i])[0])+", "+str(hurst(Is[i:1000+i],
dat[i:1000+i])[1])+", "+str(dat[i])+"\n")
out_file.close()

```

## **Dodatek C - transformacja danych do przestrzeni $[0, 1]$ oraz utworzenie zmiennej dwuwymiarowej na przestrzeni $[0, 1]^2$**

```
import numpy as np
import scipy
import matplotlib
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt
from scipy import polyfit
from numpy import zeros, floor, log10, log, mean, array, ndarray
from scipy import stats
from scipy.stats import norm, johnsonsu

lable = ["2D", "18O"]
ls = [[] for i in range(2)]
n = 0
while n < 2:
    # wczytanie danych
    c = -1
    for line in open("dane.txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 3 + n:
                    value = float(value)
                    ls[n].append(value)
            n = n + 1
    cdfs = ["norm", "johnsonsu"]
    par = [[] for i in range(2)]
    n = 0 # 1 izotop
    sample = ls[n]
    out_file = open("wynik"+str(lable[n])+".txt", "w")
    out_file.write("distribution, pvalue, Dvalue, parameters"+"\\n");
    for cdf in cdfs:
        #dopasowanie rozkladu
        parameters = eval("scipy.stats."+cdf+".fit(sample)");
        par[n].append(parameters)
        #test K-S
        D, p = scipy.stats.kstest(sample, cdf, args=parameters);
        # zapisanie parametrow do pliku
        out_file.write(cdf+"", "+str(p)+", "+str(D)+", "+str(parameters)+"\\n");
    out_file.close()

    n = 1 # 2 izotop
    sample = ls[n]
    out_file = open("wynik"+str(lable[n])+".txt", "w")
    out_file.write("distribution, pvalue, Dvalue, parameters"+"\\n");
    for cdf in cdfs:
```

```

parameters = eval("scipy.stats."+cdf+".fit(sample)");
par[n].append(parameters)
D, p = scipy.stats.kstest(sample, cdf, args=parameters);
out_file.write(cdf+","+str(p)+","+str(D)+","+str(parameters)+"\n");
out_file.close()

# transformacja do przestrzeni [0,1]
i = 0 # 1 izotop
D2 = johnsonsu.cdf(ls[i],par[i][1][0],par[i][1][1],par[i][1][2],par[i][1][3])
i = 1 # 2 izotop
O18 = johnsonsu.cdf(ls[i],par[i][1][0],par[i][1][1],par[i][1][2],par[i][1][3])

# zapisanie danych w przestrzeni [0,1] do pliku

out_file = open("johnsonsu.txt", "w")
out_file.write(lable[0]+","+lable[1]+"\\n");
for k in range(0,len (ls[0])):
    out_file.write(str(D2[k])+","+str(O18[k]+"\\n")
out_file.close()

# j.w. rozkład Gaussa
i = 0
D2 = norm.cdf(ls[i],par[i][0][0],par[i][0][1])
i = 1
O18 = norm.cdf(ls[i],par[i][0][0],par[i][0][1])

# zapisanie danych w przestrzeni [0,1] do pliku

out_file = open("normal.txt", "w")
out_file.write(lable[0]+","+lable[1]+"\\n");
for k in range(0,len (ls[0])):
    out_file.write(str(D2[k])+","+str(O18[k]+"\\n")
out_file.close()

# wykresy dopasowania jendowymiarowego
# rozkład normalny
i = 1

range = np.arange(-43, -36, 0.1)
plt.hist(ls[i], 25, normed=1, facecolor='gray', alpha=0.75)
plt.plot(range, norm.pdf(range,par[i][0][0],par[i][0][1]),'r--', linewidth=1)
plt.title(str("norm_18O"))
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.savefig("norm_18O.eps")
plt.clf()

i = 0

range = np.arange(-340, -280, 1)
plt.hist(ls[i], 25, normed=1, facecolor='gray', alpha=0.75)

```

```
plt.plot(range, norm.pdf(range,par[i][0][0],par[i][0][1]),'r--', linewidth=1)
plt.title(str("norm_D2"))
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.savefig("norm_2D.eps")
plt.clf()
# rozkład RJSU
i = 1

range = np.arange(-43, -36, 0.1)
plt.hist(ls[i], 25, normed=1, facecolor='gray', alpha=0.75)
plt.plot(range, johnsonsu.pdf(range,par[i][1][0],par[i][1][1],par[i][1][2],par[i][1][3]),'r--', linewidth=1)
plt.title(str("johnsonsu_18O"))
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.savefig("johnsonsu_18O.eps")
plt.clf()

i = 0

range = np.arange(-340, -280, 1)
plt.hist(ls[i], 25, normed=1, facecolor='gray', alpha=0.75)
plt.plot(range, johnsonsu.pdf(range,par[i][1][0],par[i][1][1],par[i][1][2],par[i][1][3]),'r--', linewidth=1)
plt.title(str("johnsonsu_2D"))
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.savefig("johnsonsu_2D.eps")
plt.clf()
```

## Dodatek D - dopasowanie do danych Kopuł Archimedesowskich

```
from __future__ import division
import scipy
import matplotlib
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt
from numpy import zeros, floor, log10, log, mean, array
from scipy import stats
from scipy.stats import kendalltau, norm, johnsonsu
import numpy as np
from scipy.integrate import quad
from scipy.optimize import fmin
import sys
from scipy.interpolate import interp1d

class Copula():

    def _integrand_debye(self,t):

        return t/(np.exp(t)-1)

    def _debye(self, alpha):

        return quad(self._integrand_debye, sys.float_info.epsilon, alpha)[0]/alpha

    def _frank_f(self, alpha):
        diff = (1-self.tau)/4.0 - (self._debye(-alpha)-1)/alpha
        return diff**2

    def __init__(self, X, Y, family):

        # rodzaje Kopuł
        copula_family = ['clayton', 'frank', 'gumbel']

        self.X = X
        self.Y = Y
        self.family = family

        # wsolczynnik korelacji tau Kendall'a
        tau = kendalltau(self.X, self.Y)[0]
        self.tau = tau
        self._get_parameter()
        self.UU = None
        self.VV = None

    def _get_parameter(self):
        # wyznaczenie parametru theta
```

```

if self.family == 'clayton':
    self.theta = 2*self.tau/(1-self.tau)

elif self.family == 'frank':
    self.theta = -fmin(self._frank_f, -5, disp=False)[0]

elif self.family == 'gumbel':
    self.theta = 1/(1-self.tau)

def generate(self, n=500):

    # generuje zmienne losowe za pomoca Kopul

    # Kopula Claytona
    if self.family == 'clayton':
        UU = np.random.uniform(size = n)
        WW = np.random.uniform(size = n)

        if self.theta < sys.float_info.epsilon :
            VV = WW
        else:
            VV = UU*(WW**(-self.theta/(1 + self.theta)) - 1 + UU**self.theta)**(-1/self.theta)

    # Kopula Franka
    elif self.family == 'frank':
        UU = np.random.uniform(size = n)
        WW = np.random.uniform(size = n)

        if abs(self.theta) > np.log(sys.float_info.max):
            VV = (UU < 0) + np.sign(self.theta)*UU
        elif abs(self.theta) > np.sqrt(sys.float_info.epsilon):
            VV = -np.log((np.exp(-self.theta*UU)*(1-WW)/WW + np.exp(-self.theta)
                        )/(1 + np.exp(-self.theta*UU)*(1-WW)/WW))/self.theta
        else:
            VV = WW

    # Kopula Gumbela
    elif self.family == 'gumbel':
        if self.theta <= 1 :
            raise ValueError('Kopula gumbela, theta =< 1')
        if self.theta < 1 + sys.float_info.epsilon:
            UU = np.random.uniform(size = n)
            VV = np.random.uniform(size = n)
        else:
            u = np.random.uniform(size = n)
            w = np.random.uniform(size = n)
            w1 = np.random.uniform(size = n)
            w2 = np.random.uniform(size = n)

            u = (u - 0.5) * np.pi
            u2 = u + np.pi/2;
            el = -np.log(w)

```



```

t = np.cos(u - u2/self.theta)/ e1
g = (np.sin(u2/self.theta)/t)**(1/self.theta)*t/np.cos(u)
sa = (-np.log(w1))**(1/self.theta)/g
sb = (-np.log(w2))**(1/self.theta)/g
UU = np.array(np.exp(-sa))
VV = np.array(np.exp(-sb))

self.UU = UU
self.VV = VV
return UU,VV

# prarmetry rozkladu normalnego oraz johnsonsu - wprowadzone recznie
parn = [[-310.0541955835962, 8.8762703751334993],[-39.350283911671916,
1.1246381464291646]]
parj = [[13.618083751773494, 6.6758417541949946, -252.80084370597098,
14.975547089604841],[12.245517426571304, 5.6447131649654079, -33.177358781096459,
1.4068449109087262]]

name = ["normal", "johnsonsu"]
name1 = ["NORMAL", "JSUD"]
lable = ["2D", "18O"]
s = 0
for s in range (0, 2):

# importuje dane w przestrzeni koncentracji
ls = [[] for i in range(2)]
n = 0
while n < 2:
    c = -1
    for line in open (str(name[s])+".txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 1 + n:
                    value = float(value)
                    ls[n].append(value)
            n = n + 1

# importuje dane w przestrzeni [0,1]
ld = [[] for i in range(2)]
n = 0
while n < 2:
    c = -1
    for line in open ("dane.txt", "r").readlines():
        c = c + 1
        if c > 0:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 3 + n:

```

```

        value = float(value)
        ld[n].append(value)
    n = n + 1

for s in range (0, 2):
    # tu zapisuje parametry Kopul
    thetaG = []

# dopasowanie kopul, generacja danych, wykresy pogladowe
# Kopula Gumbela
x = array(ls[0])
y = array(ls[1])
foo = Copula(x, y, family='gumbel')
X1, Y1 = foo.generate(2000)
if s == 0:
    X1 = norm.ppf(X1, parn[0][0], parn[0][1])
    Y1 = norm.ppf(Y1, parn[1][0], parn[1][1])
elif s == 1:
    X1 = johnsonsu.ppf(X1, parj[0][0], parj[0][1], parj[0][2], parj[0][3])
    Y1 = johnsonsu.ppf(Y1, parj[1][0], parj[1][1], parj[1][2], parj[1][3])
thetaG.append(foo.theta)
plt.scatter(ld[0], ld[1], s=20, c='r', label = "dane")
plt.scatter(X1, Y1, s=0.5, c='k', label = "symulacja")
plt.xlabel("Koncentracja Deuteru [SMOW]", fontsize=20, labelpad= -1)
plt.ylabel("Koncentracja Tlenu [SMOW]", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.legend(loc="lower right", fontsize=20)
plt.title("Kopula Gumbela", fontsize=30)
plt.savefig(str(name[s])+"_gumbelCopulaPDF.eps", format='eps', dpi = 1000)
plt.clf()

# odwrocona Kopula Gumbela
x = array(ls[0])
y = array(ls[1])
foo = Copula(x, y, family='gumbel')
X1, Y1 = foo.generate(2000)
X1 = 1-X1
Y1 = 1-Y1
if s == 0:
    X1 = norm.ppf(X1, parn[0][0], parn[0][1])
    Y1 = norm.ppf(Y1, parn[1][0], parn[1][1])
elif s == 1:
    X1 = johnsonsu.ppf(X1, parj[0][0], parj[0][1], parj[0][2], parj[0][3])
    Y1 = johnsonsu.ppf(Y1, parj[1][0], parj[1][1], parj[1][2], parj[1][3])
thetaG.append(foo.theta)
plt.scatter(ld[0], ld[1], s=20, c='r', label = "dane")
plt.scatter(X1, Y1, s=0.5, c='k', label = "symulacja")
plt.xlabel("Koncentracja Deuteru [SMOW]", fontsize=20, labelpad= -1)
plt.ylabel("Koncentracja Tlenu [SMOW]", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)

```

```
plt.legend(loc="lower right", fontsize=20)
plt.title("odw. Kopula Gumbela", fontsize=30)
plt.savefig(str(name[s])+"_REVgumbelCopulaPDF.eps", format='eps', dpi = 1000)
plt.clf()
```

# eksportuje dane do pliku (aby zrobic wykres w Originie)

```
if (s == 1):
    out_file = open("DvO_revGumbelCopula_PDF_data.txt", "w")
    out_file.write("2D, 18O \n")
    for i in range(0, len(ld[0])):
        out_file.write(str(ld[0][i])+", "+str(ld[1][i])+"\n")
    out_file.close()

    out_file = open("DvO_revGumbelCopula_PDF_sim.txt", "w")
    out_file.write("2D, 18O \n")
    for i in range(0, len(X1)):
        out_file.write(str(X1[i])+", "+str(Y1[i])+"\n")
    out_file.close()
```

# Kopula Claytona

```
x = array(ls[0])
y = array(ls[1])
foo = Copula(x, y, family='clayton')
X1, Y1 = foo.generate(2000)
if s == 0:
    X1 = norm.ppf(X1, parn[0][0], parn[0][1])
    Y1 = norm.ppf(Y1, parn[1][0], parn[1][1])
elif s == 1:
    X1 = johnsonsu.ppf(X1, parj[0][0], parj[0][1], parj[0][2], parj[0][3])
    Y1 = johnsonsu.ppf(Y1, parj[1][0], parj[1][1], parj[1][2], parj[1][3])
thetaG.append(foo.theta)
plt.scatter(ld[0], ld[1], s=20, c='r', label = "dane")
plt.scatter(X1, Y1, s=0.5, c='k', label = "symulacja")
plt.xlabel("Koncentracja Deuteru [SMOW]", fontsize=20, labelpad= -1)
plt.ylabel("Koncentracja Tlenu [SMOW]", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.legend(loc="lower right", fontsize=20)
plt.title("Kopula Claytona", fontsize=30)
plt.savefig(str(name[s])+"_claytonCopulaPDF.eps", format='eps', dpi = 1000)
plt.clf()
```

# Kopula Franka

```
x = array(ls[0])
y = array(ls[1])
foo = Copula(x, y, family='frank')
X1, Y1 = foo.generate(2000)
if s == 0:
    X1 = norm.ppf(X1, parn[0][0], parn[0][1])
    Y1 = norm.ppf(Y1, parn[1][0], parn[1][1])
elif s == 1:
    X1 = johnsonsu.ppf(X1, parj[0][0], parj[0][1], parj[0][2], parj[0][3])
```

```
Y1 = johnsonsu.ppf(Y1, parj[1][0], parj[1][1], parj[1][2], parj[1][3])
thetaG.append(foo.theta)
plt.scatter(ld[0], ld[1], s=20, c='r', label = "dane")
plt.scatter(X1, Y1, s=0.5, c='k', label = "symulacja")
plt.xlabel("Koncentracja Deuteru [SMOW]", fontsize=20, labelpad= -1)
plt.ylabel("Koncentracja Tlenu [SMOW]", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.legend(loc="lower right", fontsize=20)
plt.title("Kopula Franka", fontsize=30)
plt.savefig(str(name[s])+"_franknCopulaPDF.eps", format='eps', dpi = 1000)
plt.clf()

# parametr korelacji tau Kendall'a
print "tau"
print foo.tau
# parametry Kopul
print "theta"
print thetaG
# zaleznosc ogonowa Kopula Gumbela
print "lambda gumbel"
print 2-2**(1/thetaG[0])
print "lambda clayton"
# zaleznosc ogonowa Kopula Claytona
print 2**(-1/thetaG[2])
```



```
d = d + log(delta)
return d

def _get_parameter(self):
    # szacowanie parametrow

    if self.family == 'student':
        lista = []
        rho = sin(pi*self.tau/2)
        for i in range(1, 200):
            nu = float(i)
            lista.append([i, self.delta(X, Y, nu, rho)])
        lista = zip(*lista)
        a = lista[1].index(max(lista[1]))
        self.nu = lista[0][a]
        self.rho = rho

# generowanie danych
def generate(self, n=500):

    if self.family == 'student':
        mean = [0,0]
        cov = [[1,self.rho],[self.rho,1]]
        U = []
        V = []
        for i in range(0, n):
            ks = 0
            for i in range(0, self.nu):
                ks = ks + np.random.normal()**2
            a,b = (np.random.multivariate_normal(mean,cov,1).T)/sqrt(ks/self.nu)
            U.append(t.cdf(a,self.nu))
            V.append(t.cdf(b,self.nu))

        self.U = U
        self.V = V
        return U,V

# dalszy ciąg jak w załączniku D
parn = [[-310.0541955835962, 8.8762703751334993],[-39.350283911671916,
1.1246381464291646]]
parj = [[13.618083751773494, 6.6758417541949946, -252.80084370597098,
14.975547089604841],[12.245517426571304, 5.6447131649654079, -33.177358781096459,
1.4068449109087262]]

name = ["normal", "johnsonsu"]
name1 = ["NORMAL", "JSUD"]
lable = ["2D", "18O"]

ld = [[] for i in range(2)]
n = 0
while n < 2:
    c = -1
```

```

for line in open ("dane.txt", "r").readlines():
    c = c + 1
    if c > 0:
        a = 0
        for value in line.split(","):
            a = a + 1
            if a == 3 + n:
                value = float(value)
                ld[n].append(value)
        n = n + 1
for s in range (0, 2):

ls = [[] for i in range(2)]
n = 0
while n < 2:
    c = -1
    for line in open (str(name[s])+".txt", "r").readlines():
        c = c + 1
        if c > 1:
            a = 0
            for value in line.split(","):
                a = a + 1
                if a == 1 + n:
                    value = float(value)
                    ls[n].append(value)
            n = n + 1

for s in range (0, 2):

Y = array(ls[0])
X = array(ls[1])
foo = Copula(X, Y, "student")
print foo.nu, foo.rho
X1, Y1 = foo.generate(2000)
if s ==0:
    X1 = norm.ppf(X1, parn[0][0], parn[0][1])
    Y1 = norm.ppf(Y1, parn[1][0], parn[1][1])
elif s==1:
    X1 = johnsonsu.ppf(X1, parj[0][0], parj[0][1], parj[0][2], parj[0][3])
    Y1 = johnsonsu.ppf(Y1, parj[1][0], parj[1][1], parj[1][2], parj[1][3])
plt.scatter(ld[0], ld[1], s=20, c='r', label = "dane")
plt.scatter(X1, Y1, s=0.5, c='k', label = "symulacja")
plt.xlabel("koncentracja Deuteru [at / mg]", fontsize=20, labelpad= -1)
plt.ylabel("Koncentracja Tlenu [SMOW]", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.title("Kopula Gaussa", fontsize=30)
plt.legend(loc="lower right", fontsize=20)
plt.savefig(str(name[s])+"_StudentCopulaPDF.eps", format='eps', dpi = 1000)
plt.clf()

```

```
if (s == 1):
    out_file = open("DvO_GaussCopula_PDF_data.txt", "w")
    out_file.write("2D, 180 \n")
    for i in range(0, len(Id[0])):
        out_file.write(str(Id[0][i])+", "+str(Id[1][i])+"\n")
    out_file.close()

    out_file = open("DvO_rGaussGumbelCopula_PDF_sim.txt", "w")
    out_file.write("2D, 180 \n")
    for i in range(0, len(X1)):
        out_file.write(str(X1[i])+", "+str(Y1[i])+"\n")
    out_file.close()
```



## **Dodatek F - empiryczne wyznaczenie indeksów górnej oraz dolnej zależności ogonowej dla danych w przestrzeni $[0, 1]^2$**

```
import numpy
first = []
second = []
prog = 0.025
# wczytywanie pliku z danymi w przestrzeni  $[0,1]**2$ 
c = -1
for line in open("johnsonsu.txt", "r").readlines():
    c = c + 1
    if c > 1:
        a = 0
        for value in line.split(","):
            a = a + 1
            if a == 1:
                value = float(value)
                first.append(value)
            elif a == 2:
                value = float(value)
                second.append(value)

m = len(first)

print "lower tail dependency index"
k = 0
for i in range(0, m):
    if (first[i] <= prog) and (second[i] <= prog):
        k = k+1
# drukuje wartosc indeksu dolnej zaleznosci ogonowej
print float(k)/(m*prog)

print "upper tail dependency index"
k = 0
for i in range(0, m):
    if (first[i] >= (1-prog)) and (second[i] >= (1-prog)):
        k = k+1
#drukuj wartosc indeksu gornej zaleznosci ogonowej
print float(k)/(m*prog)
```